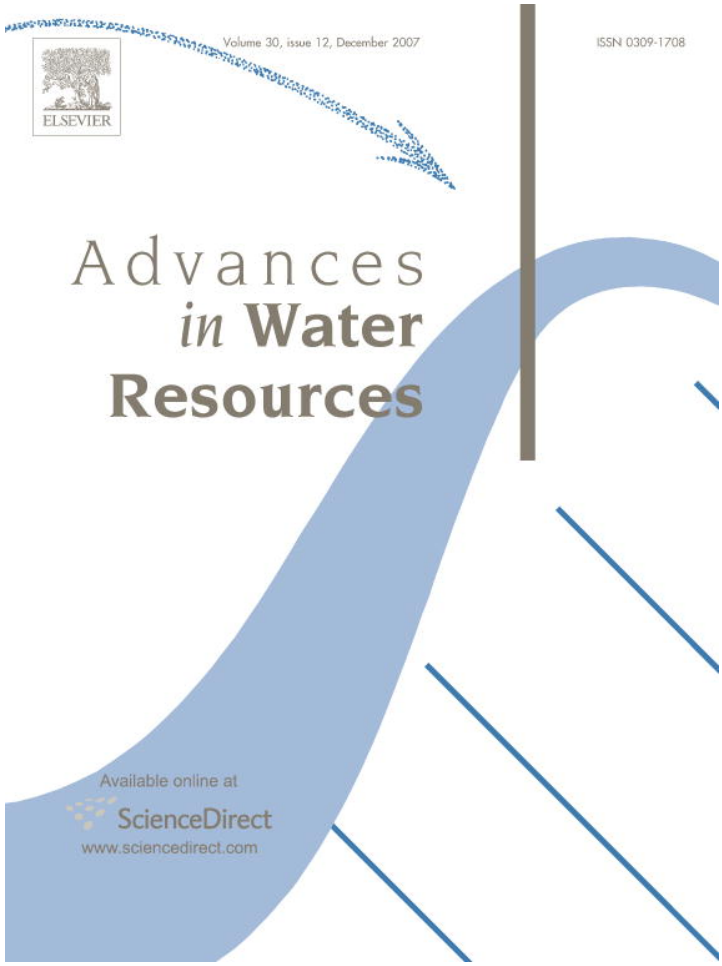

Available online at www.sciencedirect.com
ScienceDirect
Advances in Water Resources 30 (2007) 2401–2423
Advances in Water Resources
www.elsevier.com/locate/advwatres

Geospatial–temporal dependence among weekly precipitation extremes with applications to observations and climate model simulations in South America

Gabriel Kuhn [a], Shiraj Khan [a,b], Auroop R. Ganguly [a,*], Marcia L. Branstetter [c]

[a] *Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*
[b] *Department of Civil and Environmental Engineering, University of South Florida, Tampa, FL 33620, USA*
[c] *Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*

Received 4 October 2006; received in revised form 30 April 2007; accepted 6 May 2007
Available online 24 May 2007

Abstract

A quantification of the spatio-temporal dependence among precipitation extremes is important for investigating the properties of intense storms as well as flood or flash-flood related hazards. Extreme value theory has been widely applied to the hydrologic sciences and hydraulic engineering. However, rigorous approaches to quantify dependence structures among extreme values in space and time have not been reported in the literature. Previous researchers have quantified the dependence among extreme values through the concept of (pairwise bivariate) tail dependence coefficients. For estimation of the tail dependence coefficients, we apply a recently developed method [Kuhn G. On dependence and extremes. PhD thesis (Advisor: C. Klüppelberg), Munich University of Technology, 2006] which utilized the multivariate tail dependence function of a subclass of elliptical copulas. This study extends the previous approach in the context of space and time by considering pairs of spatial grids in South America and quantifying the dependence among precipitation extremes based on the time series at each spatial grid. In addition, Kendall's τ is used to estimate the pairwise copula correlation (for an elliptical copula) of precipitation between all grids in South America. The geospatial–temporal dependence measures are applied to precipitation observations from 1940 to 2005 as well as simulations from the Community Climate System Model version 3 (CCSM3) for 1940–2099. New insights are obtained regarding the spatio-temporal dependence structures for precipitation over South America both with regard to correlation as well as tail dependence.

Keywords: Geospatial; Temporal; Extremal dependence; Precipitation; Observations; Climate simulations; South America

1. Introduction

Extreme precipitation events can cause flood or flash-flood related hazards which may lead to loss of human life, property and livestock. The significance of spatial and spatio-temporal dependence among precipitation, as well as with related hydrologic or climate variables, has been increasingly recognized over the years. For example, a study of spatial and spatio-temporal dependence among precipitation extremes can be important for identifying the relationships among local intense storms and large-scale extremes or anomalies. Geospatial and temporal dependence among precipitation extremes can also be useful for hydrologic analysis and designs [2] and for understanding climatological and hydrological characteristics of the areas where the dependence is strong. However, the state of the art in hydrology and climate in the area of extremes analyses comprises mainly of applications of univariate extreme value theory.

Univariate extreme value theory has been widely used to analyze and model extremes for a single variable at a time.

* Corresponding author. Tel.: +1 865 241 1305; fax: +1 865 241 6261.
E-mail addresses: gabriel@ma.tum.de (G. Kuhn), skhan4@eng.usf.edu (S. Khan), gangulyar@ornl.gov (A.R. Ganguly), branstetterm@ornl.gov (M.L. Branstetter).

0309-1708/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2007.05.006

These approaches yield statistical estimates for $\text{Prob}(X \leqslant x + u | X > u)$, where X is the variable modeled, $x > 0$ and u is a large threshold. Another line of research has developed and utilized methodologies that model the parameters of extreme value distributions for one variable as functions of other variables, which could be time, space or other physically based covariates [3–8]. These approaches can be broadly described as $\text{Prob}(X \leqslant x + u | \mathbf{Y} = (Y_1, \ldots, Y_d), X > u)$, where X is the variable modeled via univariate extreme value theory, $x > 0$, u is the corresponding (large) threshold, and $\mathbf{Y}$ is a d-dimensional vector of covariates upon which the extremes of X are conditioned. There is another emerging line of research focussing on co-occurrence or dependence among extreme values of multiple variables [9–11]. These approaches can be broadly described as the conditional probability of the extremes of X with a threshold u given the extremes of Y with a threshold v, i.e., $\text{Prob}(X > u | Y > v)$, or the joint probability of the extremes of both variables, i.e., $\text{Prob}(X > u, Y > v)$. These approaches have the potential to quantify interesting and high-impact hydrologic relationships, for example, the dependence among precipitation or temperature extremes in space or time.

To measure spatial dependence in a spatial random field from grid-based observations, we use, because of the high-dimensionality of the problem, a concept of pairwise bivariate dependence coefficients, i.e., one single dependence parameter is assigned to each pair of grids. To distinguish between extreme and non-extreme observations, the dependence measure in this study is defined in two ways. The first one is copula correlation, i.e., the correlation parameters of an elliptical copula, which measures dependence among just precipitation values. This copula correlation (also called correlation in this study) is used as a robust dependence measure; note that estimation of copula correlation is based on Kendall's τ. The second one is tail dependence which means dependence among precipitation extremes defined as exceedances over a certain threshold. This represents the probability that an observation at one grid exceeds a large quantile, which is called threshold, given the observation at the other grid also exceeds a quantile of the same probability level [9]. The methodological contributions of this study are twofold. First, the proposed approach represents an extension of estimating multivariate extremal dependence suggested by Kuhn [1, Chapter 3] as this method is applied to high-dimensional geospatial and geospatial-temporal data (including the reduction of measuring dependence through a dependence function to a set of pairwise bivariate dependence coefficients). Second, the methods have been designed for automated dependence quantification among large-scale geospatial–temporal data without having to tailor the analysis for individual time series within pairs of grids.

Earlier domain-specific studies either applied univariate extreme value theory to large areas by analyzing each area separately [12,13] or spatial dependence among extremes to smaller regions [14,15]. Schlather and Tawn [15] analyzed daily and annual precipitation maxima in south-west England for investigating spatial extremal dependence. Previous studies dealt with the spatial and temporal variability of precipitation extremes with applications to small areas in South America [16,17]. We did not encounter a single study focussing on the spatial dependence among precipitation extremes in South America. This study investigates the spatial and temporal variability of dependence among precipitation extremes in South America for the period 1940–2005. An extreme event is defined here as an event exceeding 95%-quantile threshold. Previous studies have used high quantiles of empirical distributions as thresholds [18,19]. This study compares the correlation and tail dependence as functions of one- and two-dimensional distance (longitudinal and latitudinal distances) from observations and simulations from the Community Climate System Model version 3 (CCSM3) over South America for the period 1940–2005. The spatial and temporal correlation and tail dependence between one specific grid in the Amazon basin, Brazilian Highlands, Mato Grasso Plateau, and São Paulo, with all other grids in South America, are estimated as well as compared for the observed and simulated data. The temporal variability is investigated from 1965–2005 by considering two overlapping time windows, i.e., 1965–1990 and 1980–2005. Predictive scenarios are generated for the correlation and tail dependence by utilizing CCSM3 model simulations from 2060–2099.

Section 2 describes both observed and simulated data in South America and discusses a procedure for choosing the best data out of daily, weekly maxima, and weekly maxima residuals, which can satisfy the assumptions of extreme value analysis. Weekly maxima residuals are obtained by subtracting the long-term mean of weekly maxima of a particular week, i.e., mean of maximum weekly precipitation across the same week for all years used in the analysis, from weekly maxima of the same week. These datasets are compared based on the temporal dependence in the data using both autocorrelations and a comparison of the arrival times of exceedances over 95%-quantile thresholds with a *homogeneous Poisson process*. A simple statistic for the latter is also described in Section 2. Section 3 outlines methods for estimating the spatial dependence from the observed and simulated data using Kendall's τ for correlation, inverse ranks for extreme dependence structure, and tail dependence for extreme dependence measure. Section 4 describes the results obtained using the above mentioned methods for the observed and simulated data. The important insights, impacts of this study, and future research directions are discussed in Section 5.

2. Data availability and preparation

2.1. Data description

The daily precipitation datasets in South America are utilized in this study which are available in a gridded form of $1^0 \times 1^0$ for the period between January 1940 and June 2005 [20]. The data were collected at 7900 stations spread

over Brazil, Venezuela, North Argentina, Paraguay, Uruguay, Suriname and French Guiana. The daily precipitation at each grid was calculated by averaging daily precipitation at all stations within a radius of 0.75 times the grid spacing. The complete description of the data used in this study is available in [20]. At all grids in South America, the total number of observations used here are shown in Fig. 1. Recently, Goswami et al. [21] investigated trends of extreme rainfall events in India by analyzing daily gridded rainfall data at $1^0 \times 1^0$ resolution based on 1803 stations.

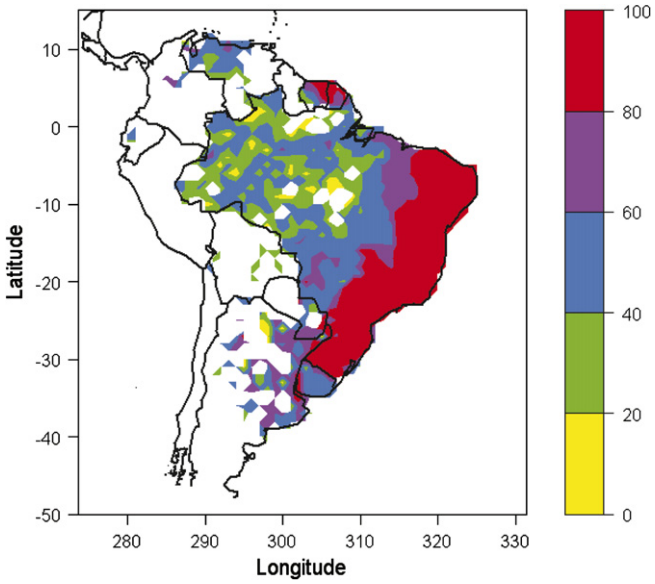

Fig. 1. Total percentage of data, given as *100*(number of observations)/(total number of possible observations from 1940–2005 available for the analysis)*.

This study also analyzes the model simulation results from January 1940 to December 2099 obtained from CCSM3 which is described by Collins et al. [22]. CCSM3 is a fully coupled climate model which includes components that represent the land, atmosphere, ocean, and sea ice. The model configuration uses a T85 grid for the land and atmosphere, which is approximately 1.4 degrees latitude and longitude, and approximately 1 degree resolution in the ocean and sea ice components. Daily precipitation results from one ensemble member of the A2 IPCC climate change scenario simulation are used for this analysis. The model simulation data obtained from Program for Climate Model Diagnosis and Intercomparison (PCMDI) includes the simulation years from 2000 to 2100. A number of other studies used similar CCSM precipitation data: Branstetter and Erickson [23] investigated the effects of precipitation characteristics on streamflow in CCSM2, Meehl et al. [24] examined precipitation intensity in CCSM3, and Boville et al. [25] described precipitation processes in CCSM3.

2.2. Temporal dependence in data

In order to estimate the spatial dependence among extremes using statistical techniques, we first need independent and identically distributed (IID) observations (or at least observations with only *weak* dependence). The assumption of identically distributed data is violated if the data contain long-term trend or seasonality whereas the temporal dependence in the data violates the assumption

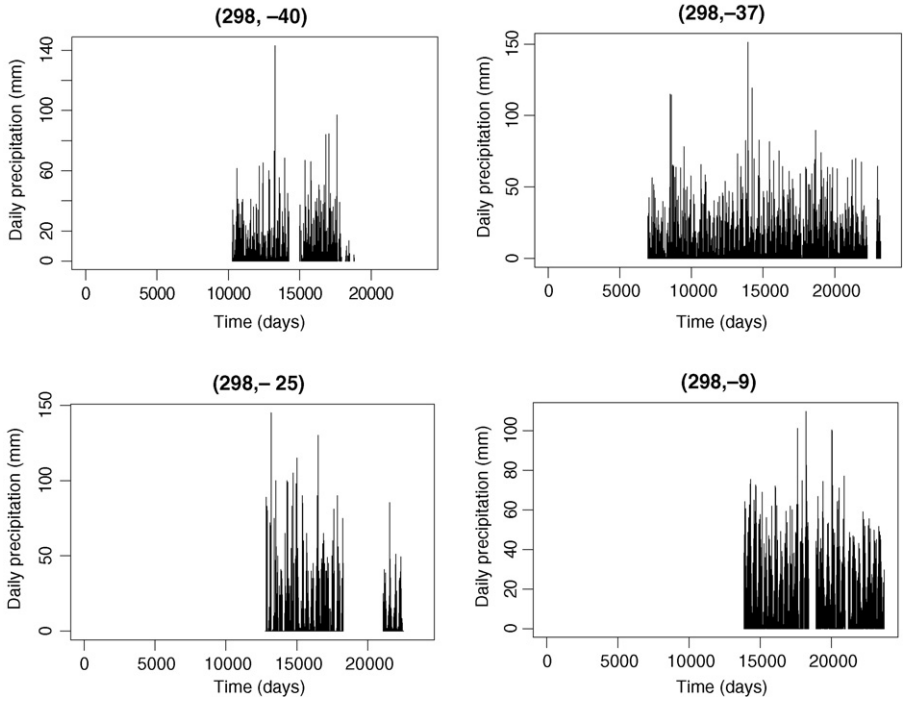

Fig. 2. Daily time series for four grids whose locations are given in terms of (*longitude*, *latitude*). The total number of observations at (298, −40), (298, −37), (298, −25), (298, −9) are 7004, 15641, 6940, and 9024, respectively.

of independent data [26]. We investigate the daily data to check if the *IID* assumptions are valid. We choose four arbitrary grids (Fig. 2), whose locations are given in terms of (*longitude*, *latitude*) as (298, −40), (298, −37), (298, −25), and (298, −9), for preliminary analyses and present initial insights into the interpretability of the results. Autocorrelation plots show that there is a significant temporal dependence in the daily data (Fig. 3). Schlather and Tawn [15] analyzed daily and annual precipitation maxima in southwest England and stated that caution should be exercised while analyzing the extremes of daily data because of temporal dependence although the daily data does contain additional information. Therefore, we do not consider the daily data for our analysis. Next, we aggregate daily data into weekly maxima, shown for four grids in Fig. 4, to check if the temporal dependence can be reduced by this procedure. But weekly maxima data do not show any improvements in the temporal dependence (Fig. 5). So, we rule out the possibility of analyzing weekly data in this study.

Next, we use a brute force method described by Gaines and Denny [26] to generate weekly maxima residuals by subtracting the long-term mean of weekly maxima of a particular week, i.e., mean of maximum weekly precipitation across the same week for all years used in the analysis, from weekly maxima of the same week. Let $X_{y,w}^{(i,j)}$ be the maximum precipitation at grid (i,j) in week w of year y. The mean of the weekly maxima of week w considering all years y (ignoring missing values) is computed as $M_w^{(i,j)} = (\sharp\{y : X_{y,w}^{(i,j)} \text{ not missing}\})^{-1} \sum_{y=1940}^{2005} X_{y,w}^{(i,j)}$, $w = 1, \ldots, 52$. Then, the residuals of weekly precipitation can be computed as $\varepsilon_{y,w}^{(i,j)} = X_{y,w}^{(i,j)} - M_w^{(i,j)}$, i.e., the precipitation value at any given week in any specific year is reduced by the mean of that week over all available years. By abbreviating the time (y, w) by t, the residuals can be represented as $\varepsilon_t^{(i,j)}$. Autocorrelation plots from weekly maxima residuals indicate that the temporal dependence in this data is significantly reduced as compared to daily and weekly maxima (Figs. 3, 5, and 6). While the simple data transformation utilized here removes the weekly mean, an added transformation is also utilized to normalize by the weekly standard deviation, i.e, $\varepsilon_{y,w}^{(i,j)} = (X_{y,w}^{(i,j)} - M_w^{(i,j)})/S_w^{(i,j)}$, where $S_w^{(i,j)}$ is the standard deviation of week w considering all years for grid (i,j). In general, the requirement is to generate high-quality models of rainfall time series at each spatial location, such that the residuals do not contain any significant temporal structure, whether in the mean or in the variance. While these areas are left for future research, we note that even the relatively simple weekly mean removal transformation reduces the autocorrelation in the time series significantly (Fig. 6a). A simple normalization by the standard deviation does not indicate any significant reduction in the autocorrelation (Fig. 6b).

The removal of the weekly means may imply that the residuals and their extremes lose the typical physical interpretability associated with them because seasonal fluctuations are removed from the extremes. In this study, the extremes of the residuals represent those extremes that are large enough compared to what could be typically expected for a given week for a given spatial location. In this sense, the selected thresholds are local, not only in space but also in time. In this study, the maxima in precipitation, and hence the thresholds, are significantly larger

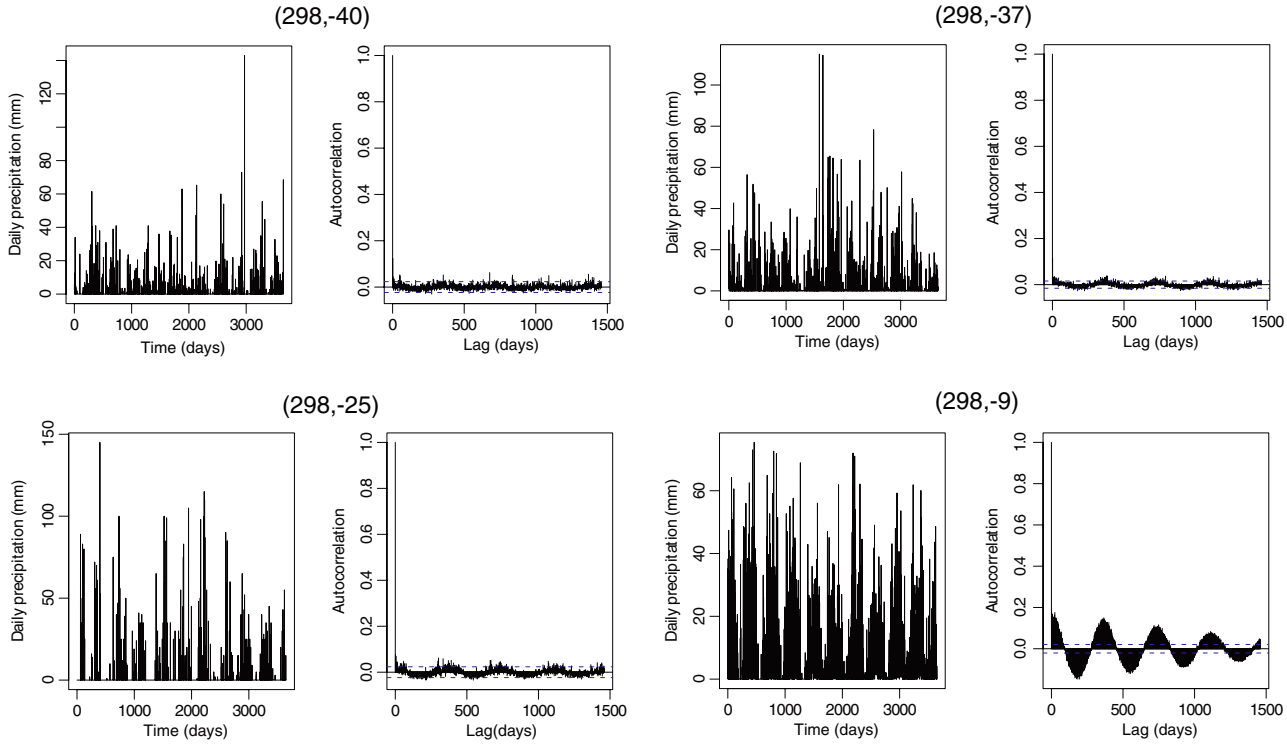

Fig. 3. Daily time series showing 10 years of data for four grids and their autocorrelation plots.

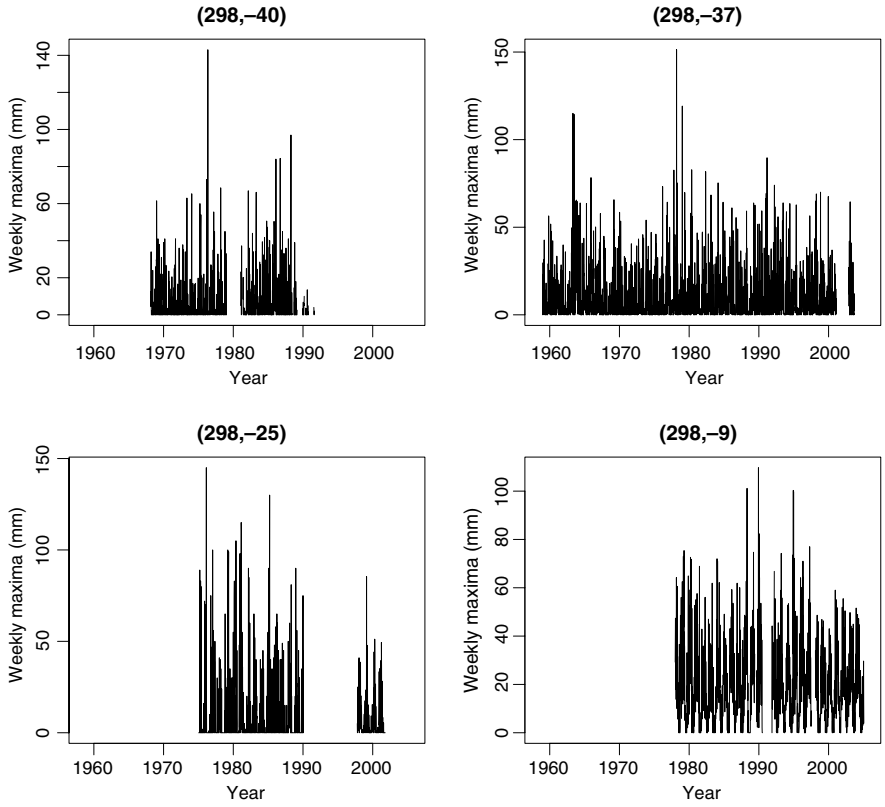

Fig. 4. Weekly maxima data for four grids.

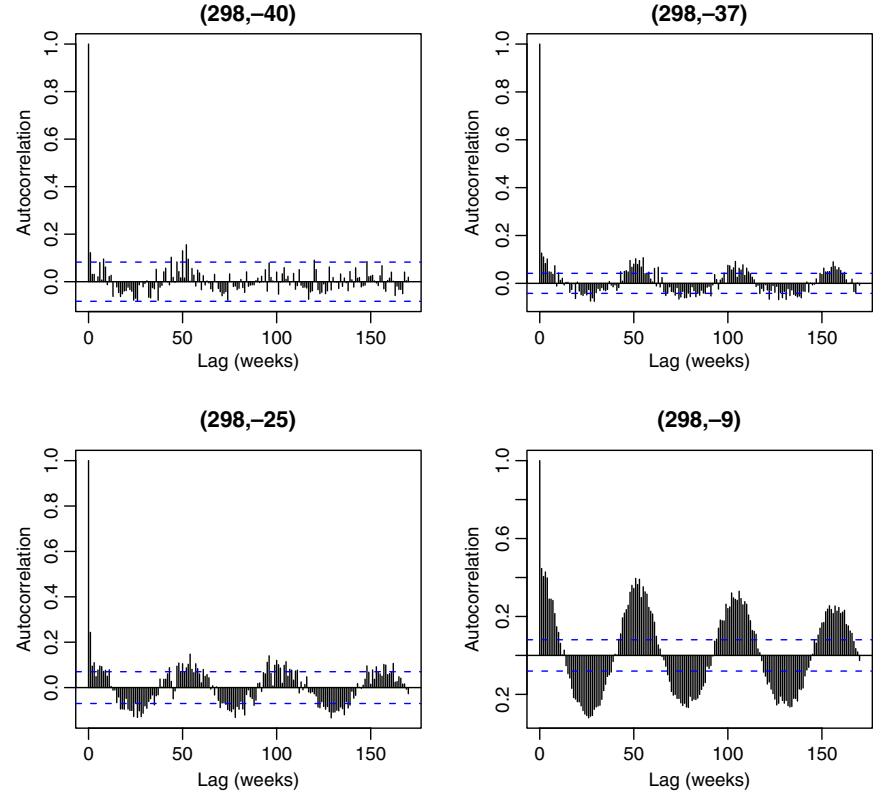

Fig. 5. Weekly maxima data: autocorrelation plots for four grids.

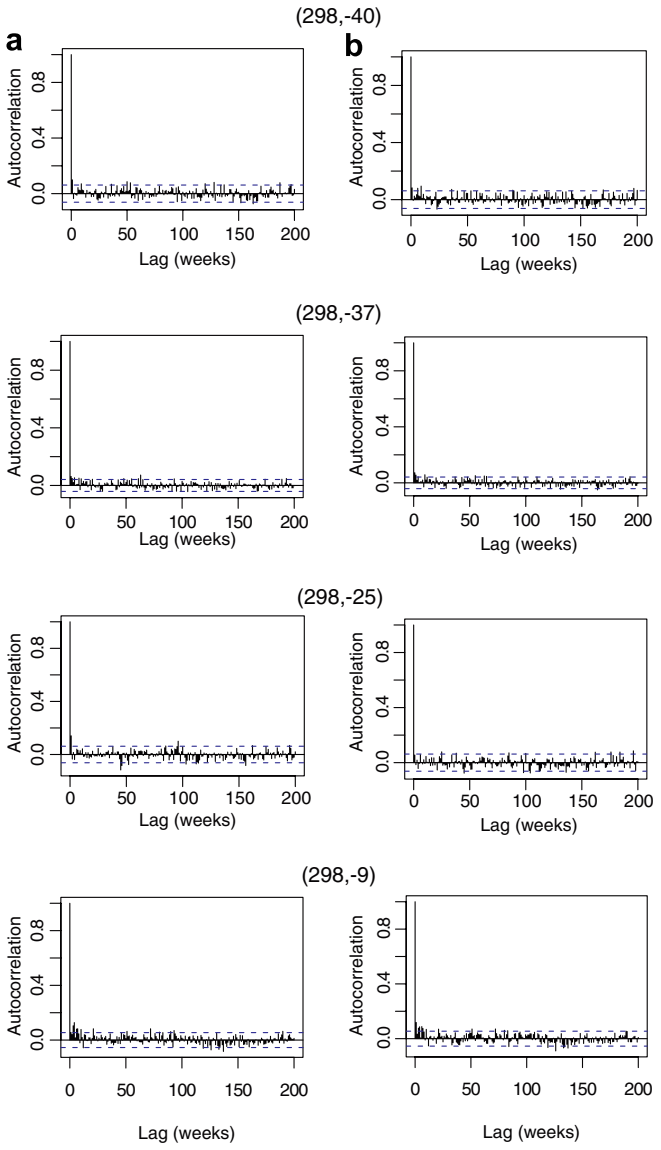

Fig. 6. Weekly maxima residuals: autocorrelation plots for four grids. Weekly maxima residuals are generated from (a) $\varepsilon_{y,w}^{(i,j)} = X_{y,w}^{(i,j)} - M_w^{(i,j)}$ and (b) $\varepsilon_{y,w}^{(i,j)} = (X_{y,w}^{(i,j)} - M_w^{(i,j)})/S_w^{(i,j)}$. Since we do not observe any significant changes from (b) if compared with (a), we utilize (a) to generate weekly maxima residuals for this study.

compared to the seasonal variations, thus even the loss of physical interpretation is minimal. Thus, the reduction by the weekly mean is a worthwhile tradeoff given the gain in applicability of extreme value theory, as evidenced by the reduction in autocorrelation while moving from weekly maxima to weekly maxima residuals (Figs. 5 and 6). Therefore, this study utilizes weekly maxima residuals generated from $\varepsilon_{y,w}^{(i,j)} = X_{y,w}^{(i,j)} - M_w^{(i,j)}$ for the analysis. Kunkel et al. [27] analyzed weekly precipitation totals generated from daily data for understanding the spatial and temporal characteristics of extremes in the Midwest of US whereas Yates et al. [28] investigated the spatial and temporal dependencies of the climate variables by utilizing weekly precipitation residuals.

2.3. Check for IID assumptions of threshold exceedances

The focus of this study is to investigate the spatial dependence among precipitation extremes. In order to estimate the spatial dependence among extremes, we first check the *IID* assumptions of exceedances over a large threshold defined as 95%-quantile. If the exceedances over a large threshold are *IID*, their inter-arrival times follow a *homogeneous Poisson process* and the excesses over a threshold follow a *generalized Pareto distribution* (GPD) [3, Chapter 7].

This study checks if the inter-arrival times of threshold exceedances follow a *homogeneous Poisson process*. According to the definition of *homogeneous Poisson process*, the inter-arrival times of threshold exceedances should be exponentially distributed. In order to compare the distribution of the inter-arrival times with the exponential distribution, we use the goodness-of-fit statistic D_{SP}, which is suggested by Michael [29] and based on the *stabilized probability plot*. Let k inter-arrival times of threshold exceedances be $t_1, \ldots, t_k$ and $t_1 < \cdots < t_k$ be an ordered sample drawn from an exponential distribution, whose cumulative distribution function is given as $F_0(t, \lambda) = 1 - \mathrm{e}^{-\lambda t}$ for $t \geqslant 0$. The stabilized probability plot consisting of coordinates (r_i, s_i) is given as

$$r_i = \frac{2}{\pi} \arcsin \sqrt{\frac{1}{k}\left(i - \frac{1}{2}\right)},$$

$$s_i = \frac{2}{\pi} \arcsin \sqrt{\frac{1}{\hat{\lambda}} F_0(t_i, \hat{\lambda})},$$

where $\hat{\lambda}$ is the maximum likelihood estimator of λ under an exponential distribution. The stabilized plot indicates the departures of the theoretical values by the deviations of the plotted points from a line joining $(0,0)$ and $(1,1)$ [30]. Michael [29] mentioned that the variances of plotted points are approximately equal from the stabilized plot and this property motivated the definition of a goodness-of-fit statistic, D_{SP}, given as

$$D_{\mathrm{SP}} = \max_{i=1,\ldots,k} |r_i - s_i|,$$

which removes the subjectivity in the interpretation of stabilized plots by measuring the maximum deviation of the plotted points from their theoretical values. D_{SP} is analogous to the standard Kolmogorov–Smirnov statistic but it is more powerful [29,31]. We use D_{SP} to measure the maximum deviation of the inter-arrival times of threshold exceedances from an exponential distribution. D_{SP} can be compared with critical values D_{SP}^*, which is obtained as some sample quantile recorded from m number of samples of sample size n [30]. For normal, logistic, Cauchy, and double exponential distributions, Coles [30] calculated D_{SP}^* as 95%-quantile of 10000 samples of size 10, 25, and 40 data points. Since this study analyzes 1053 grids, we consider 1000 samples for the calculation of D_{SP}^* in the interests of computational tractability. We generate 1000 independent

samples of sample size n and consider each sample to calculate D'_{SP} from the inter-arrival times of exceedances over 95%-quantile threshold. Next, we consider 1000 D'_{SP}s to compute D^*_{SP}, which is given as the 95%-quantile. In order to compare D_{SP} and D^*_{SP}, we define a simple measure as

$$\overline{D}_{SP} = \frac{D_{SP}}{D^*_{SP}}.$$

If $\overline{D}_{SP} > 1$, we reject with 95% confidence that the inter-arrival times of threshold exceedances are exponentially distributed, hence the inter-arrival times of threshold exceedances do not follow a *homogeneous Poisson process*. If $\overline{D}_{SP} \leqslant 1$, we do not reject the assumption that the inter-arrival times of threshold exceedances follow a *homogeneous Poisson process*. This study uses $\overline{D}_{SP}$ at all 1053 grids because it can be easily computed, plotted and visualized in space for comparisons.

Weekly maxima residuals and the corresponding QQ-plots of the inter-arrival times of threshold exceedances against the exponential distribution are shown in Fig. 7 for the four grids. It is observed that the inter-arrival times match the exponential distribution quite well except for the grid $(298, -25)$ where there is still evidence of significant clustering. We note that clustering in this case means that large precipitation events in one week are more likely to be followed by large events in the consecutive week. $\overline{D}_{SP}$ for daily and weekly maxima residuals from observations, and for weekly maxima residuals from CCSM3 simulations are shown in Fig. 8. We observe significant improvements in $\overline{D}_{SP}$ from weekly maxima residuals as compared to daily data. From weekly maxima residuals, the major parts of South America indicate $\overline{D}_{SP} \leqslant 1$ which also emphasizes our point of using weekly maxima residuals for the analysis. While $\overline{D}_{SP}$ described here is rather rigorous, the *IID* constraints may need to be relaxed for real-world applications from pragmatic considerations. However, the fact that the value of this statistic exceeds unity at some parts of South America does indicate a need for caution while interpreting the results.

As described earlier, the presence of temporal dependence in the precipitation time series implies that the extremal insights need to be interpreted carefully. We note that the problem of completely extracting the temporal dependence from precipitation time series has been a challenging

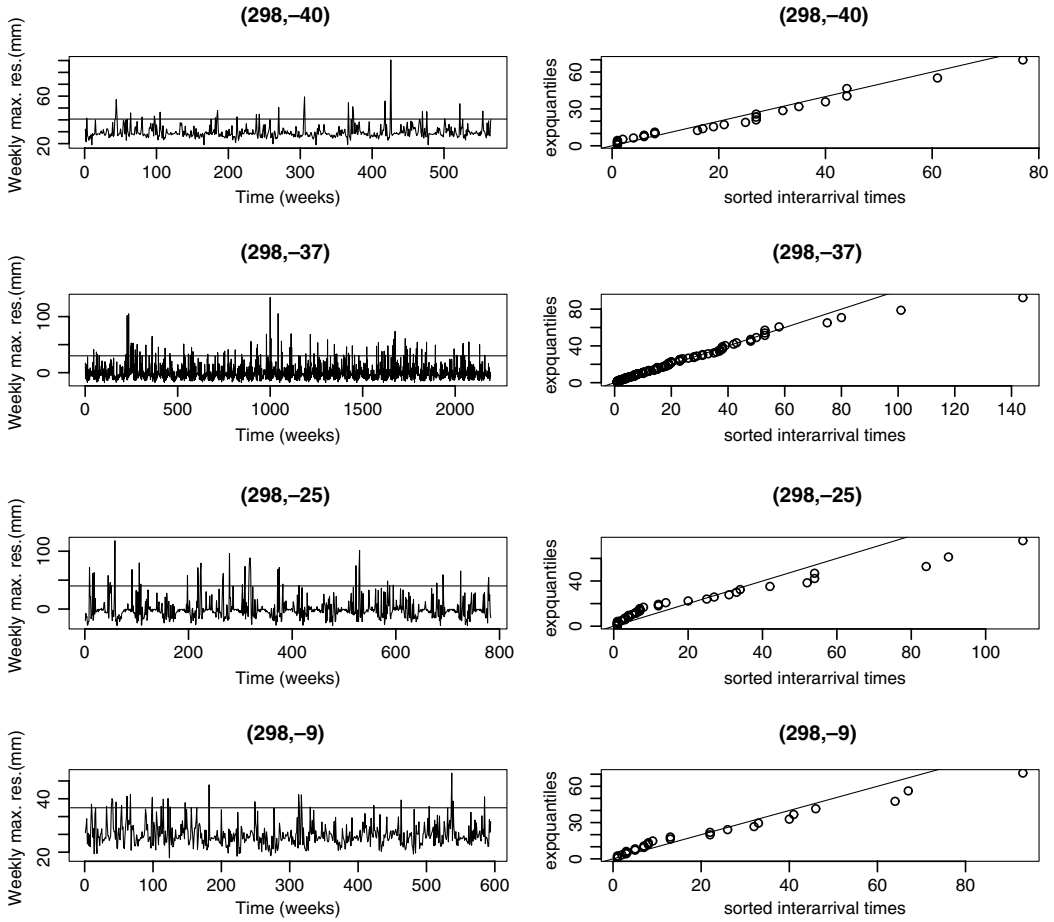

Fig. 7. Left column: Weekly maxima residuals for grids $(298, -40)$, $(298, -37)$, $(298, -25)$ and $(298, -9)$ with thresholds (shown as a horizontal line) chosen as 95%-quantile. This study defines an extreme event if it exceeds 95%-quantile threshold. Right column: Corresponding QQ-plots of inter-arrival times between threshold exceedances vs exponential distribution. The inter-arrival times between threshold exceedances is a sequence of time intervals at which extremes occur.

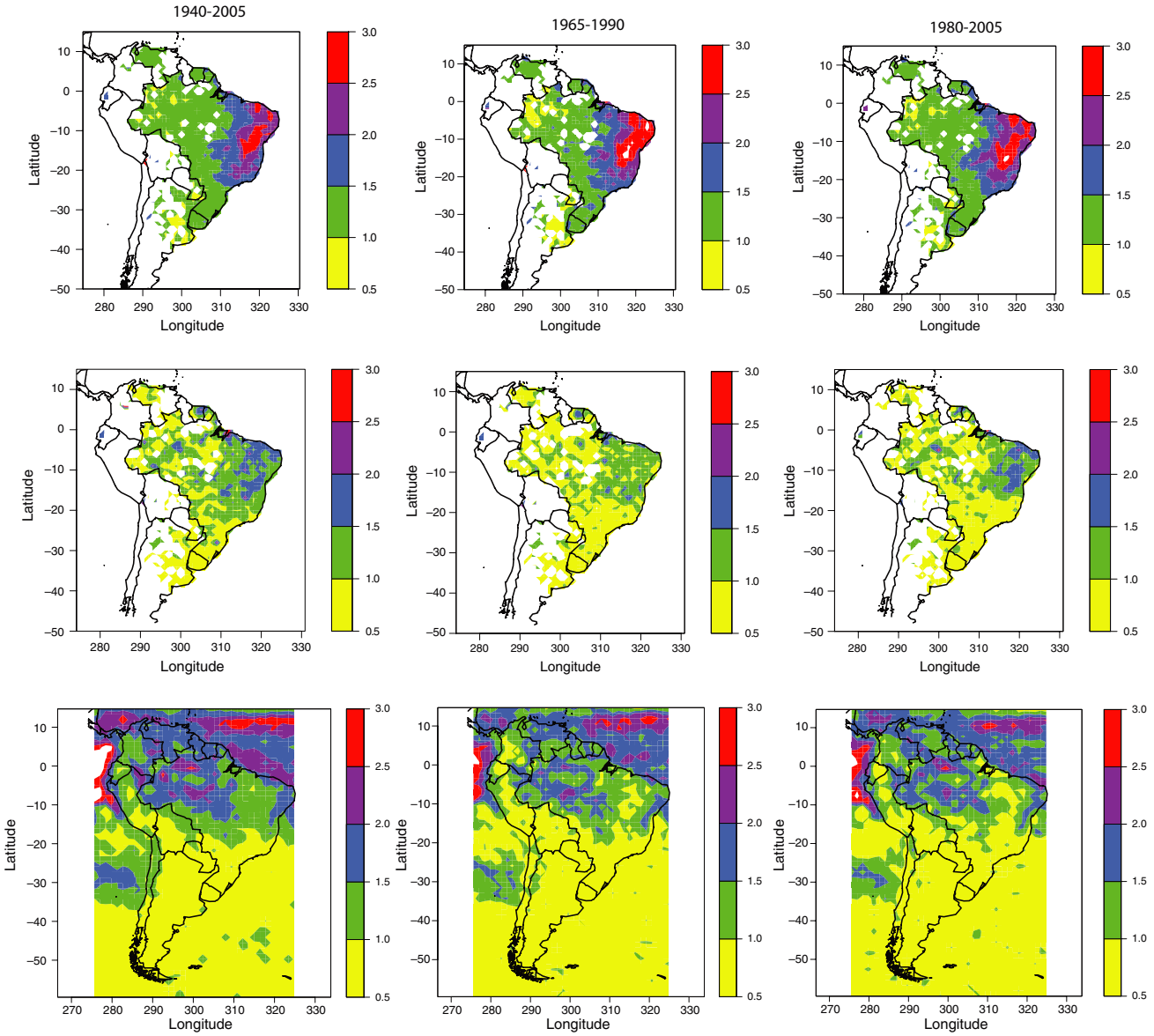

Fig. 8. Goodness-of-fit statistic, $\overline{D}_{SP}$, of inter-arrival times of exceedances over 95%-quantile against exponential distribution for the whole South America. Top row: Observed daily data, Middle row: Observed weekly maxima residuals, and Bottom row: Simulated weekly maxima residuals from the CCSM3 climate model.

problem for hydrologists and statisticians. Previous studies have dealt with the problem of temporal dependence in the context of extreme value analysis [4–6,32–36]. A few of these studies modeled non-stationarity in time for univariate analysis by defining the GPD parameters as functions of time [4–6]. This procedure is not straightforward for the estimation of spatial dependence, since the standard approach requires spatial data which are independent in time. Since this study analyzes 1053 grids, it is not feasible to find the best model corresponding to each grid which can satisfy the *IID* assumptions for rigorous applications of extreme value theory. For this reason, this study utilizes a simple model, even though more rigorous models and powerful testing methods for model selection can be used on a grid-by-grid basis.

3. Spatial dependence analysis

3.1. Kendall's τ, copula, and elliptical copula

Kendall's τ: A rank-based dependence measure, such as Kendall's τ, can be utilized to estimate the dependence among precipitation residuals. If $(X, Y) \sim F$ is a pair of random variables with distribution F and $(\widetilde{X}, \widetilde{Y}) \sim F$ is independent of (X, Y) (and with the same distribution F), Kendall's τ between X and Y is defined as

$$\tau := P((X - \widetilde{X})(Y - \widetilde{Y}) > 0) - P((X - \widetilde{X})(Y - \widetilde{Y}) < 0).$$

Given an *IID* sample $(X_1, Y_1), \ldots, (X_n, Y_n)$, the empirical estimator of Kendall's τ is

$$\hat{\tau} := \frac{2}{n(n-1)} \sum_{1 \leqslant i < j \leqslant n} \operatorname{sign}((X_i - X_j)(Y_i - Y_j)). \quad (1)$$

Kendall's τ is more robust compared to linear correlation in the following ways: (a) It does exist even if the first or second moments of (X, Y) do not exist; (b) A single outlier, or a few outliers, cannot lead to misleading values for the estimator; and (c) It is invariant under strictly increasing transformations (as opposed to being invariant only under strictly increasing linear transformations). Detailed information about Kendall's τ is given in [37].

Copula: Let $\boldsymbol{X} = (X_1, \ldots, X_d)^{\mathrm{T}}$ be a random vector with distribution function F and continuous margins $F_1, \ldots, F_d$. By Sklar's theorem [38,39], there exists a unique copula $C : [0,1]^d \to [0,1]$ such that $F(\boldsymbol{x}) = C(F_1(x_1), \ldots, F_d(x_d))$, $\boldsymbol{x} = (x_1, \ldots, x_d)^{\mathrm{T}} \in \mathbb{R}^d$. The copula is a d-dimensional distribution function with standard uniform margins. According to Sklar's theorem, a distribution can always be separated into its margins and copula. The copula can be interpreted as the dependence structure of the distribution. Therefore, the margins can be ignored in order to analyze the dependence structure of a random vector.

Elliptical copula: A random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ has an *elliptical distribution*, if

$$\boldsymbol{X} \overset{d}{=} \boldsymbol{\mu} + G\mathbf{A}\boldsymbol{U}^{(d)}, \quad (2)$$

where $\boldsymbol{\mu}$ is a deterministic d-dimensional vector; $G > 0$ is a random variable; $\mathbf{A}$ is a deterministic $d \times d$ matrix with $\mathbf{A}\mathbf{A}^{\mathrm{T}} := \boldsymbol{\Sigma} = (\sigma_{ij})_{1 \leqslant i,j \leqslant d}$ and $\operatorname{rank}(\boldsymbol{\Sigma}) \leqslant d$; $\boldsymbol{U}^{(d)}$ is a d-dimensional random vector uniformly distributed on the unit hyper-sphere $\mathscr{S}_d := \{\boldsymbol{z} \in \mathbb{R}^d : \boldsymbol{z}^{\mathrm{T}}\boldsymbol{z} = 1\}$; and $\boldsymbol{U}^{(d)}$ is independent of G. An *elliptical copula* is defined as the copula of an elliptically distributed random vector. Due to the invariance of copulas under strictly increasing transformations, an elliptical copula can be uniquely characterized by the correlation matrix $\mathbf{R} = (\rho_{ij})_{1 \leqslant i,j \leqslant d} := [\operatorname{diag}(\boldsymbol{\Sigma})]^{-1/2} \boldsymbol{\Sigma} [\operatorname{diag}(\boldsymbol{\Sigma})]^{-1/2}$ and the distribution of the generating variable G (unique up to some positive constant). This correlation is the linear correlation only if the entire distribution (and not just the copula) is elliptical and if the first and second moments exist. Hence, if the copula is elliptical but the distribution is not elliptical, this correlation measure does not necessarily convey any information about the linear correlation of the random vector. Therefore, this dependence measure may be referred to as a *copula correlation*. In case of an elliptical copula (with $\operatorname{Prob}(G = 0) = 0$), copula correlation and Kendall's τ are related as

$$\rho_{ij} = \sin(\tau_{ij}\pi/2), \quad 1 \leqslant i, j \leqslant d, \quad (3)$$

where $(\rho_{ij})_{1 \leqslant i,j \leqslant d}$ is the copula correlation matrix [40]. A correction term can be introduced if $\operatorname{Prob}(G = 0) \neq 0$ [40]. Using Eqs. (1) and (3), the copula correlation can be estimated using Kendall's τ by retaining all the robustness properties of Kendall's τ estimator, i.e.,

$$\hat{\rho}_{ij} := \sin(\hat{\tau}_{ij}\pi/2), \quad 1 \leqslant i, j \leqslant d. \quad (4)$$

There are several other (standard) dependence measures, such as Pearson linear correlation or the rank-based Spearman's rho. The Pearson linear correlation has the drawback of being only a linear dependence measure, and it is inappropriate and misleading in the case of multivariate log-normal distribution [41]. Also, this estimator is only defined for finite first and second moments. Both Kendall's τ and Spearman's rho are rank based, and therefore copula dependence measures make them invariant under strictly increasing transformations of the margins. Both of them usually show similar behavior and estimation values. However, for an elliptical copula a relation like Eq. (3) does not hold in the case of Spearman's rho [40]. The behavior of several correlation estimators is given in [42]. Consistency and asymptotic normality of the Kendall's τ based correlation estimator $\hat{\rho} = \sin(\hat{\tau}\pi/2)$ are derived in [1, Theorem 4.4.9]. The confidence bounds for $\hat{\rho}$ are not derived here. The estimator of the asymptotic covariance matrix of $\hat{\rho}$ is defined in [1, Definition 4.4.10] and its consistency and asymptotic normality are shown in [1, Theorem 4.4.11].

3.2. Tail dependence function of an elliptical copula for extreme dependence measure

A standard approach to measure dependence among extremes is given through *multivariate extreme value theory* (MEVT) and the componentwise block maxima approach therein. We give a short overview of the probabilistic theory of dependence among extremes and point to Resnick [43], Hsing et al. [44], and Einmahl et al. [45] for more details. Given an *IID* sequence of d-dimensional random vectors $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,d})^{\mathrm{T}}$, $1 \leqslant i \leqslant n$, the componentwise block maxima are defined as

$$\boldsymbol{M}_1 = (X_{1,1}, \ldots, X_{1,d})^{\mathrm{T}} \quad \text{and}$$

$$\boldsymbol{M}_n = (M_{n,1}, \ldots, M_{n,d})^{\mathrm{T}} = \left(\bigvee_{i=1}^{n} X_{i,1}, \ldots, \bigvee_{i=1}^{n} X_{i,d} \right)^{\mathrm{T}}, \quad n > 1,$$

where '$\bigvee$' denotes the maximum among the arguments. Assume there exists sequences $a_{n,j} > 0$ and $b_{n,j} \in \mathbb{R}$, $1 \leqslant j \leqslant d$, such that

$$P\left(a_{n,1}^{-1}(M_{n,j} - b_{n,j}) \leqslant x_j, 1 \leqslant j \leqslant d\right) \overset{n \to \infty}{\to} H(x_1, \ldots, x_d) =: H(\boldsymbol{x}),$$
$$x \in \mathbb{R}^d,$$

where H is a non-degenerate d-dimensional distribution function. Then, any possible limit H is called *multivariate extreme value distribution function* (MEVDF). It can easily be shown that a cumulative distribution function (CDF) H is a MEVDF if and only if the marginals H_j, $1 \leqslant j \leqslant d$, are one-dimensional extreme value CDFs [46] and the copula C of H is an extreme value copula, i.e., it satisfies [38, Section 6.2]

$$C^t(u_1, \ldots, u_d) = C(u_1^t, \ldots, u_d^t), (u_1, \ldots, u_d) \in [0,1]^d, \quad t > 0.$$

It is also known that any extreme copula can be written in terms of the *Pickands representation* [43, Section 5.4], i.e.,

$$C(u_1,\ldots,u_d)=\exp\left\{\int_{\mathscr{S}_d}\left(\bigwedge_{j=1}^{d} w_j(\ln u_j)\right)\mu(\mathrm{d}\boldsymbol{w})\right\},\quad (u_1,\ldots,u_d)\in[0,1]^d, \tag{5}$$

where '$\bigwedge$' denotes the minimum among the arguments and μ is a finite measure on the unit simplex $\mathscr{S}_d=\{\boldsymbol{s}\geqslant\boldsymbol{0}:\sum_{j=1}^{d}s_j=1\}$ satisfying

$$\int_{\mathscr{S}_d} w_\mu(\mathrm{d}\boldsymbol{w})=1,\quad j=1,\ldots,d.$$

Changing the variable $\boldsymbol{w}$ in the integral in the Pickands' representation as shown in Eq. (5), the extreme copula can be described in infinitely many different but equivalent forms, e.g., Einmahl et al. [45] adopts the following representation for the bivariate case:

$$C(u_1,u_2)=\exp\left\{\int_0^{\pi/2}\left(\frac{\ln u_1}{1\vee\cot\theta}\wedge\frac{\ln u_2}{1\vee\tan\theta}\right)\Phi(\mathrm{d}\theta)\right\},\quad u_1,u_2\in[0,1],$$

where Φ is a finite measure (called *spectral measure*) on $[0,\pi/2]$ satisfying

$$\int_0^{\pi/2}(1\vee\tan\theta)\Phi(\mathrm{d}\theta)=\int_0^{\pi/2}(1\vee\cot\theta)\Phi(\mathrm{d}\theta)=1.$$

One aim of this study is to estimate and visualize the spatial dependence among precipitation extremes. Since we consider a very high dimensional data set with 1053 dimensions, dependence representations through some $(d-1)$-dimensional measure μ or a spectral measure is difficult to handle both for estimation and visualization. Therefore, we focus on the pairwise bivariate dependence concept called *tail dependence coefficient*. Pairwise bivariate dependence parameters lead to a dependence matrix like correlation but here in terms of extremal dependence. For the notion of tail dependence coefficient, we first need another equivalent representation of extremal dependence, i.e., the *tail dependence function* which is defined for a random vector $\boldsymbol{X}$ with CDF F and margins F_j, $1\leqslant j\leqslant d$, through the limit as

$$\lambda^{\boldsymbol{X}}(x_1,\ldots,x_d):=\lim_{t\to 0}\frac{1}{t}P(1-F_1(X_1)\leqslant tx_1,\ldots,1-F_d(X_d)\leqslant tx_d),$$

with $x_1,\ldots,x_d\geqslant 0$ and if the limit exists. The (pairwise bivariate) tail dependence coefficient between components X_i and X_j, $1\leqslant i,j\leqslant d$, is defined as

$$\begin{aligned}\lambda^{\boldsymbol{X}}_{i,j}&:=\lambda^{\boldsymbol{X}}(\infty,\ldots,\infty,1,\infty,\ldots,\infty,1,\infty,\ldots,\infty)\\&=\lim_{t\to 0}\frac{1}{t}P(1-F_i(X_i)\leqslant t,1-F_j(X_j)\leqslant t)\\&=\lim_{t\to 0}P\left(X_j\geqslant F_j^{\leftarrow}(1-t)\middle|X_i\geqslant F_i^{\leftarrow}(1-t)\right),\end{aligned} \tag{6}$$

where $F_j^{\leftarrow}$ denotes the generalized inverse of F_j and if the limit exists. Note that the last line in Eq. (6) represents the usual definition of the tail dependence coefficient, e.g., see [38] (the only difference is that we use '$\geqslant$' whereas Joe's definition uses '$>$'). When $d=2$, the tail dependence function, i.e., $\lambda^{\boldsymbol{X}}(x,y)$, can be estimated non-parametrically through bivariate extreme value theory (see [45] and references therein). In addition, parametric models for the tail dependence function have been suggested and estimated [47,48,3]. The non-parametric and parametric estimation of tail dependence functions have been typically investigated for $d=2$ only, even though both approaches can be theoretically applicable for $d>2$. For higher dimensions, the non-parametric estimation of tail dependence is given in [44].

One intuitive interpretation of the tail dependence coefficient can be given in terms of t-year return levels. If X_i and X_j are two locations with continuous and strictly increasing CDFs and t-year return levels z_i and z_j, respectively, i.e., $P(X_i>z_i)=P(X_j>z_j)=1/t$, then

$$P(X_i>z_i,X_j>z_j)=\frac{1}{t}P(X_j>z_j|X_i>z_i)\sim\frac{\lambda^{\boldsymbol{X}}_{i,j}}{t},\quad \text{as } t\to\infty,$$

i.e., the event that both locations X_i and X_j exceed their t-year return levels simultaneously is approximately a $(t/\lambda^{\boldsymbol{X}}_{i,j})$-year event, if t is sufficiently large. Note that the notion $a(t)\sim b(t)$ as $t\to\infty$ means $\lim_{t\to\infty}[a(t)/b(t)]=1$. This representation through tail dependence coefficients for a random vector $\boldsymbol{X}$ leads to a *tail dependence matrix* given as

$$\boldsymbol{\Lambda}^{\boldsymbol{X}}:=\begin{pmatrix}\lambda^{\boldsymbol{X}}_{1,1}&\cdots&\lambda^{\boldsymbol{X}}_{1,d}\\ \lambda^{\boldsymbol{X}}_{2,1}&\cdots&\lambda^{\boldsymbol{X}}_{2,d}\\ \vdots&\ddots&\vdots\\ \lambda^{\boldsymbol{X}}_{d,1}&\cdots&\lambda^{\boldsymbol{X}}_{d,d}\end{pmatrix}.$$

We next focus on the estimation of tail dependence coefficients. Given an *IID* sample $(X_{1,i},X_{1,j}),\ldots,(X_{n,i},X_{n,j})$, the standard empirical estimator for the bivariate tail dependence function between the two components X_i and X_j with CDFs F_i and F_j, respectively, is defined as

$$\widehat{\lambda}^{\mathrm{emp}}_{i,j}(x_i,x_j;k):=\frac{1}{k}\sum_{l=1}^{n}\boldsymbol{I}\left(1-\widehat{F}_i(X_{l,i})\leqslant\frac{k}{n}x_i,1-\widehat{F}_j(X_{l,j})\leqslant\frac{k}{n}x_j\right), \tag{7}$$

where $\widehat{F}_i$ and $\widehat{F}_j$ denote the empirical distributions of $\{X_{l,i}\}_{1\leqslant l\leqslant n}$ and $\{X_{l,j}\}_{1\leqslant l\leqslant n}$, respectively, and for $n\to\infty$, $k=k(n)\to\infty$ and $k/n\to 0$. Hence, the standard empirical estimator for the tail dependence coefficient between X_i and X_j is given through $\widehat{\lambda}^{\mathrm{emp}}_{i,j}(1,1;k)$. But estimation in extreme value theory is always tricky and results are unstable and sensitive with respect to k. Also, the *optimal* choice of k is still an unresolved problem. Peng [49] described one method to choose an optimum k and this method was extended by Einmahl et al. [50] recently.

This study uses an another approach to improve the estimation of the tail dependence coefficient through a semi-parametric model for the tail dependence function (see [1] for more details). We assume that the tail dependence function (not the complete distribution) of (X_i,X_j) is the same as the tail dependence function of an elliptical copula which is the copula of an elliptical random vector $G_{i,j}\mathbf{A}_{i,j}\boldsymbol{U}^{(2)}$ with

copula correlation coefficient $\rho_{i,j}$ (see Eq. (2) with explanations below) and the generating variate $G_{i,j}$ is regularly varying with index $\alpha_{i,j}$, i.e., $P(G_{i,j} < tx)/P(G_{i,j} > t) \overset{t\to\infty}{\to} x^{-\alpha_{i,j}}$. Given these assumptions, the tail dependence function as shown in [1] (Chapter 3) is given as

$$\begin{aligned}\lambda^{\boldsymbol{X}}_{i,j}(x,y) &= \lambda^{\mathrm{El}}(x,y;\alpha_{i,j},\rho_{i,j}) \\ &= \left(x\int_{g_{i,j}((x/y)^{1/\alpha_{i,j}})}^{\pi/2} (\cos\phi)^{\alpha_{i,j}}\,\mathrm{d}\phi + y\int_{g_{i,j}((x/y)^{-1/\alpha_{i,j}})}^{\pi/2} (\cos\phi)^{\alpha_{i,j}}\,\mathrm{d}\phi \right) \\ &\quad \times \left(\int_{-\pi/2}^{\pi/2} (\cos\phi)^{\alpha_{i,j}}\,\mathrm{d}\phi \right)^{-1}, \end{aligned} \tag{8}$$

where $g_{i,j}(t) = \arctan((t-\rho_{i,j})/\sqrt{1-\rho_{i,j}^2})$ and $\rho_{i,j}$ is the copula correlation coefficient between X_i and X_j.

Some remarks on the tail dependence function of an elliptical copula are given as follows: (a) The multivariate representation of the tail dependence function $\lambda^{\mathrm{El}}(\boldsymbol{x},\mathbf{R},\alpha)$ of an elliptical copula is also shown in [1] and only depends on the arguments $\boldsymbol{x} = (x_1,\ldots,x_d) \geqslant \mathbf{0}$, the copula correlation matrix $\mathbf{R}$ and the regular variation index α; (b) In case of a Gaussian copula, $G^2 \sim \chi^2_d$ and, hence $P(G > tx)/P(t) \overset{t\to\infty}{\to} 0 = x^{-\infty}$ for all $x > 0$ and, therefore, $\lambda^{\boldsymbol{X}}_{i,j}(x_i,x_j) = 0$ for all $x_i, x_j \geqslant 0$ and if $\rho_{i,j} < 1$. In such a case where $\lambda \equiv 0$, the random vector (and its components) are called *asymptotically independent*, otherwise they are called *asymptotically dependent*; (c) An example of an elliptical copula with $0 < \alpha < \infty$ is the t_ν-copula, i.e., the copula of the (t_ν)-distribution of the random vector $\mathbf{A}\boldsymbol{Z}/\sqrt{\chi^2_\nu/\nu}$, where $\mathbf{A}$ is some deterministic matrix, $\boldsymbol{Z}$ is a standard normal vector, $\nu > 0$ and χ^2_ν is a χ^2_ν-distributed random variable. If $\mathbf{R}$ be the corresponding copula correlation matrix, the tail dependence function is $\lambda^{\mathrm{El}}(\cdot,\mathbf{R},\nu)$ for this t_ν-copula; (d) We do not require observation being heavy-tailed or multivariate regularly varying (see [43] for the latter term). We only require that the tail dependence function is the same as derived from an elliptical copula where the corresponding elliptical distribution satisfies some conditions on regular variation.

The whole procedure for the estimation of the tail dependence function can be briefly described as follows: (i) Estimate copula correlation, $\hat{\rho}_{i,j}$ defined in Eq. (4), using all the data; (ii) Estimate the empirical tail dependence function, $\hat{\lambda}^{\mathrm{emp}}_{i,j}(x,y;k)$ as shown in Eq. (7), for some (x,y) using all the data above the 95%-quantile threshold, and (iii) $\alpha_{i,j}$ is estimated through inversion of the (theoretical) tail dependence function given in Eq. (8). These steps are repeated for several pairs (x,y) leading to set of (partial) $\alpha_{i,j}$ estimators. The final $\alpha_{i,j}$-estimator is then defined as a weighted mean of the partial $\alpha_{i,j}$-estimators with some weight function w and denoted by $\hat{\alpha}_{i,j}(k,w)$. Note that not all combinations of positive (x,y) can be used for estimation since the tail dependence function is homogeneous and not invertible everywhere. A choice of different (x,y) and a weight function w are given in [1]. Finally, the tail dependence estimator is defined as

$$\hat{\lambda}^{\boldsymbol{X}}_{i,j}(k,w) := \lambda^{\mathrm{El}}(1,1;\hat{\alpha}_{i,j}(k,w),\hat{\rho}_{i,j}).$$

This particular estimator is much smoother than the standard empirical estimator since more points are considered for its estimation. Note that this estimator is mainly determined by the empirical estimators $\widehat{\lambda}^{\mathrm{emp}}(x,y;k)$ and the theoretical tail dependence function λ^{El} is only used to interpolate between them for different (x,y). The asymptotic properties of this estimator have been described and compared with the standard empirical estimator in [1].

Often, estimation of (theoretically) small tail dependence has some bias, i.e., tail dependence is overestimated. In this case, a large bias may occur because convergence towards the limit is slow and occurs very far out in a region where no data are available. For example, in the analysis done later, it may be reasonable to assume no tail dependence between locations which are more than 2000 km apart even though some of these locations do show some dependence based on our estimates.

We provide a couple of reasons in the defence of our approach based on the tail dependence function of an elliptical copula. First, elliptical copulas provide a large class of copulas which are very flexible and easy to handle (also in high dimensions). Other classes of copulas suffer especially in higher dimensions since they have a very restricted set of dependence parameters, like the class of *Archimedean copulas* [39]. Second, in order to improve (and smooth) the standard empirical estimator through interpolation, we need some parametric tail dependence functions. Of course, one can choose commonly used parametric tail dependence function from another copula, e.g., *Gumbel* copula, or just use the standard empirical tail dependence estimator. Comparisons between tail dependence estimators are given by Frahm et al. [51].

3.3. Parallel computational implementations

For this study, we develop an R code base for estimating dependence among extreme values from time series between pairs of spatial grids. The dependence is geospatial in the sense that grid pairs in space are investigated, while the temporal aspect comes in since we use time series observations at each grid and because multiple time windows are utilized. The number of grids, i.e., 1053, and the temporal windows, result in more than half a million pairs for which the dependence measure needs to be computed. In addition, the computations need to be performed for both observations and climate model simulations. However, the dependence measures between each pair are computed once, and subsequently these measures can be utilized to display and visualize the results in various ways such that they make sense to hydrologic and climate modelers. The leadership class computing facilities available at the Oak Ridge National Laboratory are utilized for the computations of the dependence measures. The problem is parallelized with relative ease and hence the computations are split into smaller chunks. Each computation job is programmed in such a way that they can leverage hundreds of CPUs

(multiple processes with multiple CPUs per process). The significant computational challenge presents interesting research challenges which can be investigated in depth in follow-up studies. However, for this specific study, the computational challenges preclude a detailed uncertainty analysis, as well as the dependence among other observations and climate variables. Future efforts will need to investigate these issues in much more depth.

4. Results

4.1. Correlation and tail dependence among four grids

4.1.1. Correlation for dependence among precipitation residuals

The overall correlation in the weekly maxima precipitation residuals are investigated first. The pairwise plots of the four grids shown in Fig. 9 suggest that linear correlation may not be an appropriate choice for a dependence measure. Thus, for reasons described earlier, *Kendall's* τ has been utilized to estimate dependence among precipitation residuals. Correlation can be estimated by plugging *Kendall's* τ in Eq. (4). Correlation between $(298,-40)$ and $(298,-37)$ is high because these grids are closely located whereas there is no correlation among other grids as shown in Table 1.

4.1.2. Inverse ranks for extreme dependence structure

The inverse ranks method is used to explore and visualize the dependence structure among precipitation extremes between four grids. Given the time series of the residuals $\varepsilon_1^{(i,j)},\ldots,\varepsilon_n^{(i,j)}$, let $\varepsilon_{(1,n)}^{(i,j)} > \varepsilon_{(2,n)}^{(i,j)} > \cdots > \varepsilon_{(n,n)}^{(i,j)}$ denote the order statistics. Further, let $R_t^{(i,j)}$ be the rank corresponding to observation $\varepsilon_t^{(i,j)}$, i.e., $R_t^{(i,j)} = k$ if and only if $\varepsilon_t^{(i,j)} = \varepsilon_{(k,n)}^{(i,j)}$. Then, based on Example 4.3 in [44], $(1/R_t^{(i_1,j_1)}, 1/R_t^{(i_2,j_3)})_{1\leqslant t\leqslant n}$ are plotted for all pairs $(i_1,j_1),(i_2,j_2) \in \{(298,-40), (298,-37),(298,-25)$ and $(298,-9)\}$, (truncated by points close to $(1,1)$ for easy viewing) as shown in Fig. 10. Such plots give an insight in the complete extreme dependence structure between datasets, i.e., points close to the axes indicate independent extremes whereas points away from the axes show the dependence structure of extremes. For grids $(298,-40)$ and $(298,-37)$ most of the observations lie close to the axes

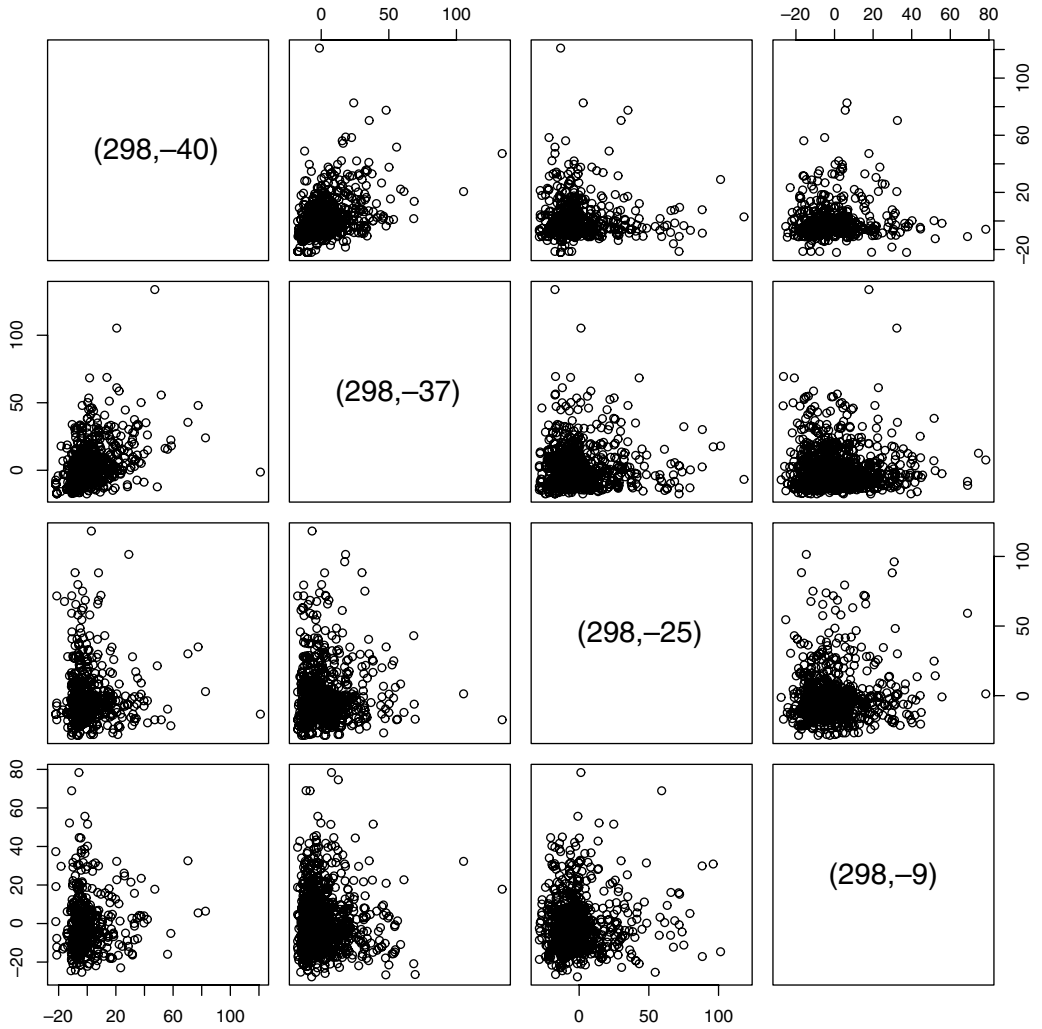

Fig. 9. Scatter plots of weekly maxima precipitation residuals in *mm* for four grids.

Table 1
Correlation based on Kendall's τ and tail dependence given in [ ] between the four grids

Grid	(298, −40)	(298, −37)	(298, −25)	(298, −9)
(298, −40)	1.0 [1.0]	0.518 [0.268]	0.017 [0.031]	0.050 [0.045]
(298, −37)	0.518 [0.268]	1.0 [1.0]	0.059 [0.074]	0.026 [0.053]
(298, −25)	0.017 [0.031]	0.059 [0.074]	1.0 [1.0]	−0.023 [0.088]
(298, −9)	0.050 [0.045]	0.026 [0.053]	−0.023 [0.088]	1.0 [1.0]

Correlation and tail dependence between two closely located grids, i.e., (298, −40) and (298, −37), are high whereas correlations among other grids are close to zero and tail dependences among other grids seem to be overestimated.

but some observations also lie away from the axes, i.e., if a large rainfall is observed at one of these grids, it is likely in some cases to observe a large rainfall at the other grid. The extreme dependence between grid (298, −40) and (298, −25) or (298, −9) is weak, i.e., most of the extremes occur independently. Moreover, this plot shows a quite symmetric extreme dependence structure, hence it is reasonable to model this with a symmetric model like an elliptical copula described in Section 3.1.

4.1.3. Tail dependence for extreme dependence measure

The tail dependence is used for quantifying the pairwise dependence among precipitation extremes corresponding to the four grids. Table 1 shows some dependence in the extremes between nearby grids (298, −40) and (298, −37) but almost no tail dependence between the others. If we assume that tail dependence decays with spatial distance, it is probably not reasonable to assume any tail dependence between (298, −40), (298, −25), and (298, −9) due to large

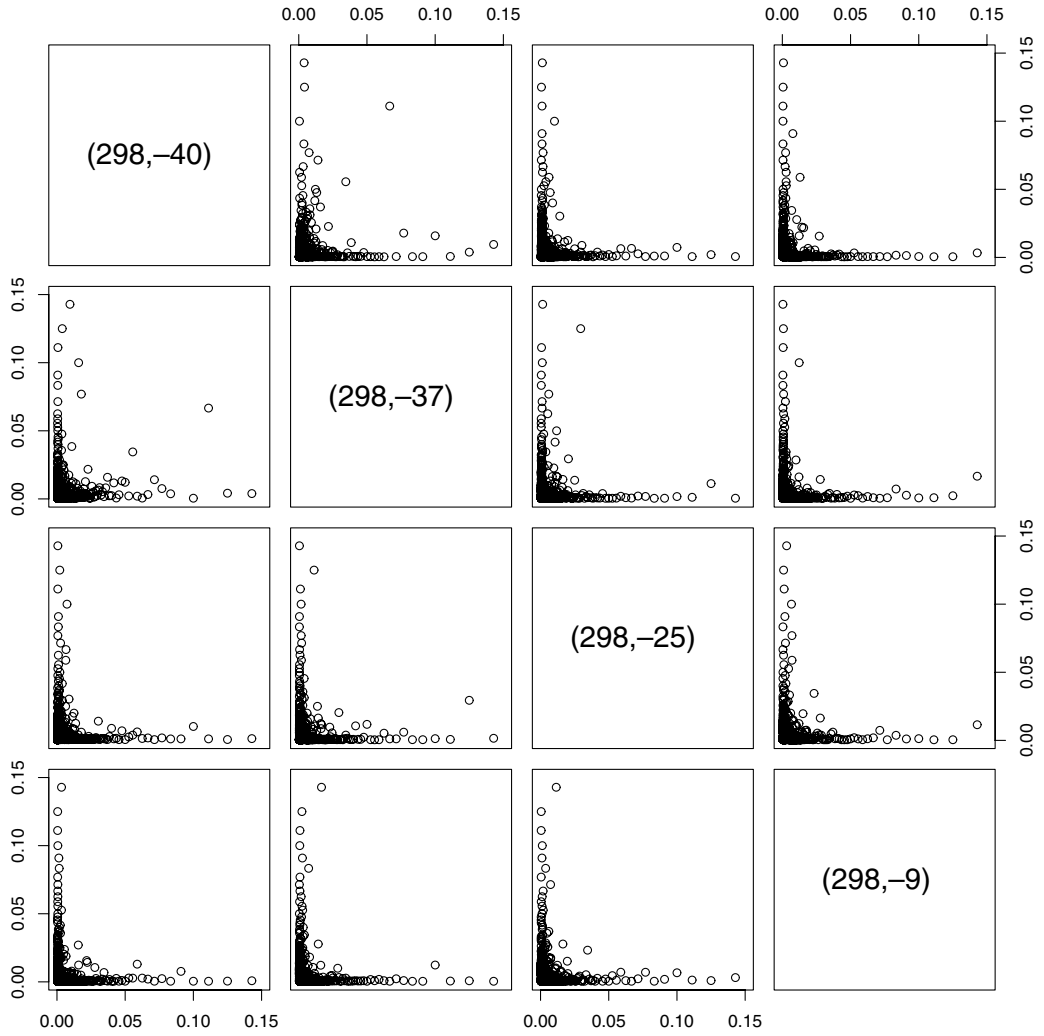

Fig. 10. Inverse ranks, i.e., $(1/R_t^{(i_1,j_1)}, 1/R_t^{(i_2,j_2)})_{1\leqslant t\leqslant n}$, plots for four grids.

distance. This indicates possible bias in the estimates. The problem of asymptotic tail independence, i.e., zero tail dependence, is not considered here. Methods for detecting asymptotic tail independence are given in [52,11].

4.2. Correlation and tail dependence with distance in the whole South America

The variability of dependence, i.e., correlation and tail dependence, is investigated with respect to one-dimensional distance, or the spatial proximity alone without regard to direction or location, for the whole South America. Using the same approach used earlier for the four grids, two 1053×1053-matrices for correlation and tail dependence are generated. Alternatively, we generate four-dimensional arrays for correlation, i.e., $\mathrm{Corr}(\mathrm{lon}_1, \mathrm{lat}_1; \mathrm{lon}_2, \mathrm{lat}_2)$, where lon and lat are longitude and latitude, respectively, and for tail dependence, i.e., $\lambda(\mathrm{lon}_1, \mathrm{lat}_1; \mathrm{lon}_2, \mathrm{lat}_2)$, which implies the correlation and tail dependence between locations with coordinates $(\mathrm{lon}_1, \mathrm{lat}_1)$ and $(\mathrm{lon}_2, \mathrm{lat}_2)$. Each pair of grids $(i,j)_a$ and $(i,j)_b$, $a \neq b \in \{1, \ldots, 1053\}$ are considered and the distance between them is calculated by using the *Haversine* formula. This generates a $(1053 \cdot 1052/2) \times 2$ matrix containing distances and their corresponding correlations and tail dependences.

Correlation and tail dependence decrease with distance (Figs. 11a and 11b). In south-west England, Schlather and Tawn [15] investigated spatial extremal dependence using

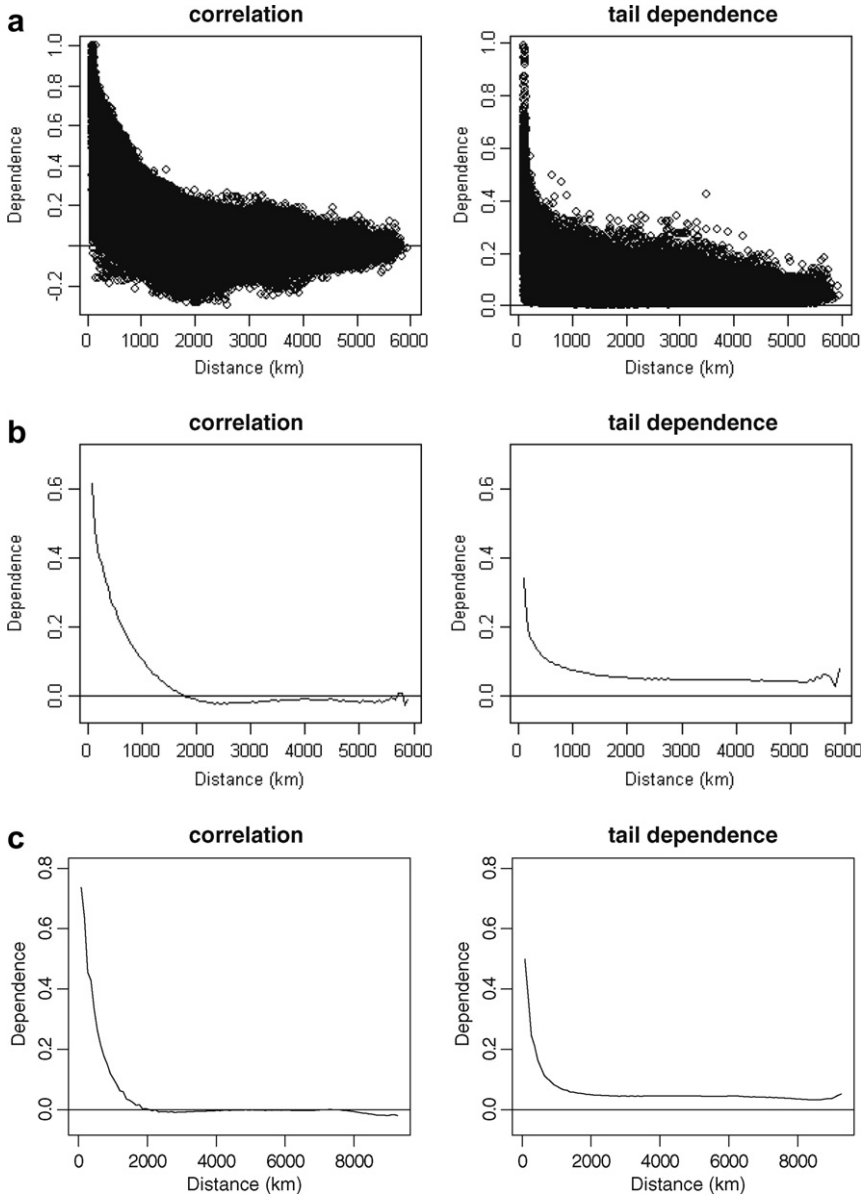

Fig. 11. Variation of dependence (correlation and tail dependence) vs. distance for 1940–2005. (a) Observed data: correlation and tail dependence; (b) Observed data: 10 km moving averages, i.e., the average dependence for distances in an interval $(d, d + 10\ \mathrm{km})$, $d = 0$ km, 10 km, 20 km, ..., and (c) CCSM3 simulations: 10 km moving averages, i.e., the average dependence for distances in an interval $(d, d + 10\ \mathrm{km})$, $d = 0$ km, 10 km, 20 km, ...

daily and annual precipitation maxima and found that it decays smoothly with distance. The maximum negative correlation is −0.2 for some points located within 6000 km but the average correlation is close to zero between all points situated beyond 1700 km. We note that the decay of spatial correlation with distance, while commonly observed for geographic data, can have exceptions due to dependence caused by topography induced atmospheric dynamics as well as due to long-range teleconnections. However, we do expect correlations to become negligible on the average with large distances, which provides the rationale for our assumption here. In general, care must be taken while applying this assumption lest valid dependencies are ignored. The possibility of some positive bias in the tail dependence estimates appears likely given that the dependence becomes relatively steady when the distance between grids is greater than 2000 km (Fig. 11b). By subtracting this (assumed) positive bias of 0.05 from the tail dependence estimates, it is observed that tail dependence is very close to zero beyond 1500 km. Analysis of simulations from CCSM3 for 1940–2005 shows higher values of correlation and tail dependence compared to that obtained from the analysis of observations (Figs. 11b and 11c). Correlation becomes zero whereas tail dependence becomes steady which suggests a possible positive bias when the distance between grids is greater than 2000 km (Fig. 11c).

Considering observed data for two time windows 1965–1990 and 1980–2005, correlation between grids located within 1700 km increases from 1965 to 2005 (Fig. 12). If the (assumed) positive bias is taken out from the tail dependence estimates for 1965–1990 and 1980–2005, an increase in tail dependence is observed between grids located at a distance less than 1500 km from 1965 to 2005. From CCSM3 simulations, the correlation and tail dependence plots for three time windows, i.e., 1965–1990, 1980–2005, and 2060–2099, are not different from that obtained for the period 1940–2005 (plots not shown here).

4.3. Correlation and tail dependence with longitudinal and latitudinal distance in the whole South America

Correlation and tail dependence with longitudinal and latitudinal distances between grids are investigated here. For each grid pair, the longitudinal and latitudinal distance are calculated by accounting for both the distance and the directionality between the two grids. This generates a $(1053 \cdot 1052/2) \times 3$ matrix containing two-dimensional distances and their corresponding correlations and tail dependences.

The spatial dependence, i.e., both correlation and tail dependence, obtained from observations displays an elliptical pattern with the longer axis along the latitudinal direction for the period 1940–2005 (Fig. 13a). This indicates additional dependence generally along the latitudinal direction, over and above the dependence caused by geographical proximity. Since latitudes demarcate climatic zones,

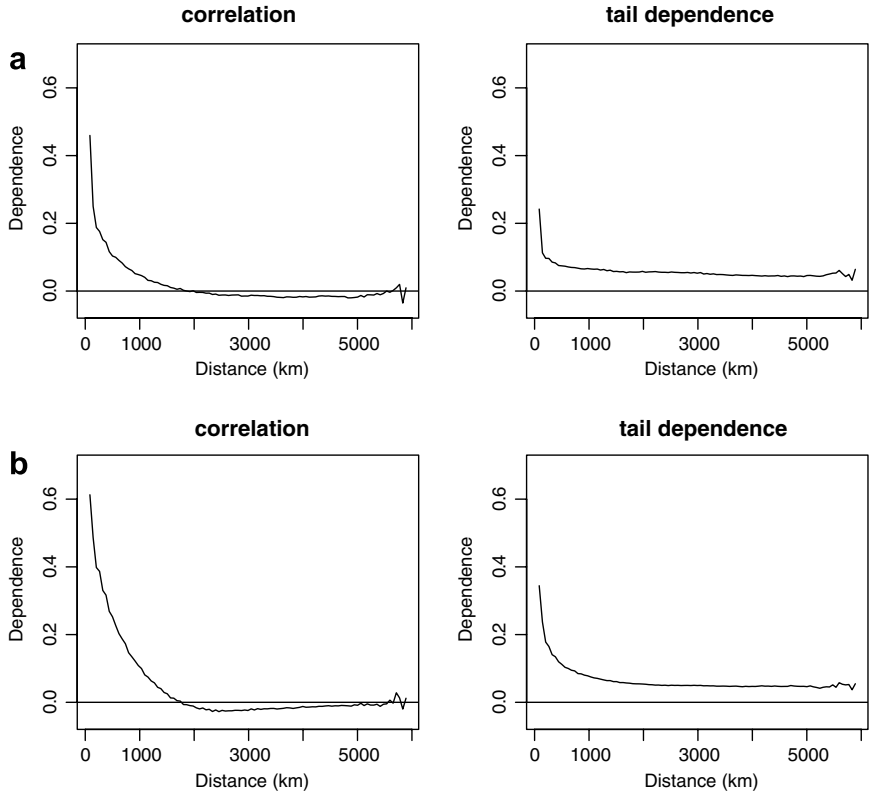

Fig. 12. Observed data: variation of dependence (correlation and tail dependence) vs. distance for (a) 1965–1990 and (b) 1980–2005. These plots show the 10 km moving averages, i.e., the average dependence for distances in an interval $(d, d + 10\text{ km})$, $d = 0$ km, 10 km, 20 km, ...

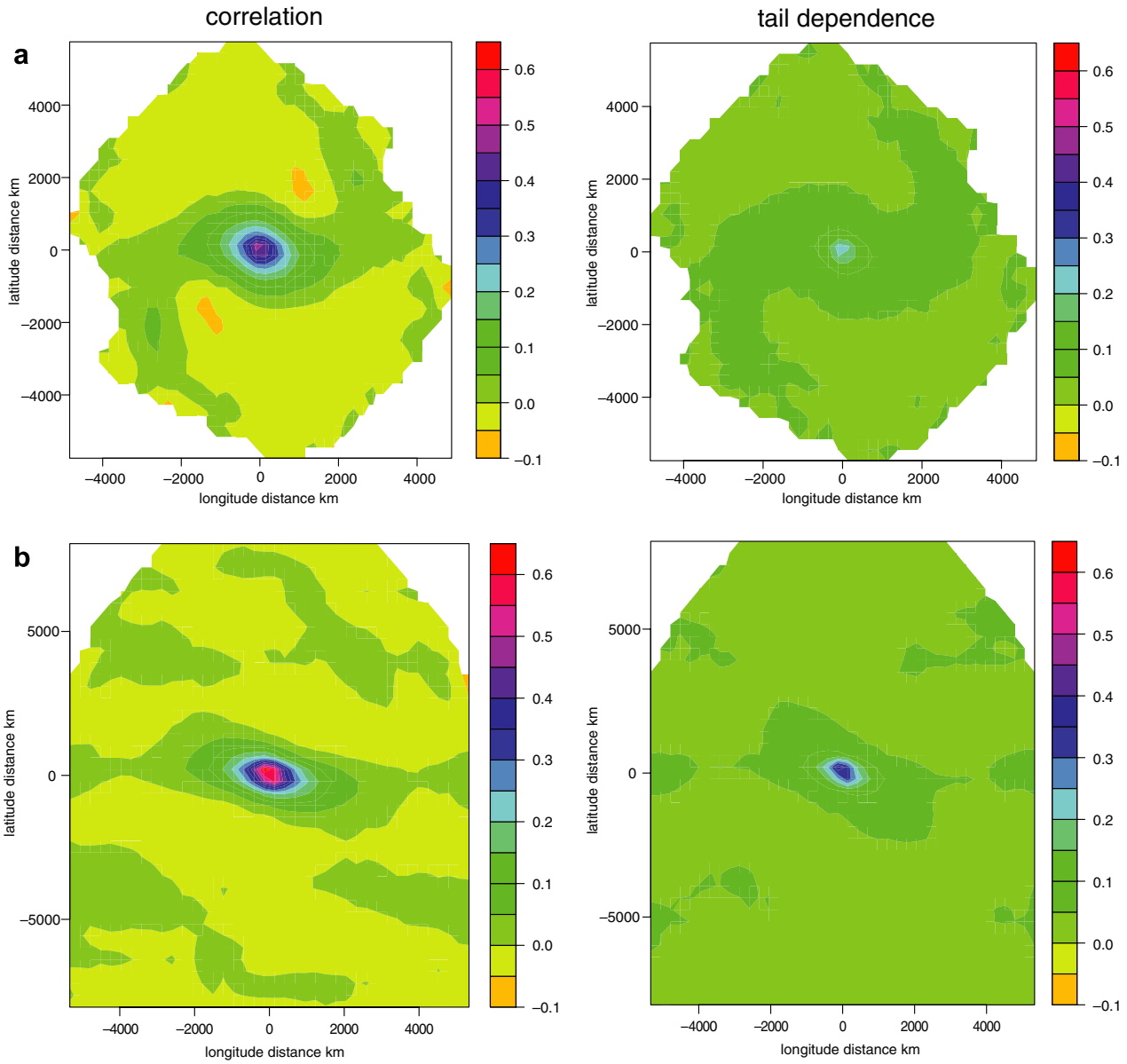

Fig. 13. Dependence (correlation and tail dependence) vs. longitude/latitude distance averages in South America for the period 1940–2005. (a) Observed data: averages of all points lying inside a square whose diagonal is given by $(d_{\text{lat}}, d_{\text{long}})$ and $(d_{\text{lat}} + 200\text{ km}, d_{\text{long}} + 200\text{ km})$, where $d_{\text{lat}} = (-6000\text{ km}, -5800\text{ km}, \ldots, 6000)$ and $d_{\text{long}} = (-5000\text{ km}, -4800\text{ km}, \ldots, 5000)$ and (b) CCSM3 simulations: averages of all points lying inside a square whose diagonal is given by $(d_{\text{lat}}, d_{\text{long}})$ and $(d_{\text{lat}} + 200\text{ km}, d_{\text{long}} + 200\text{ km})$, where $d_{\text{lat}} = (-7500\text{ km}, -7300\text{ km}, \ldots, 7500)$ and $d_{\text{long}} = (-5000\text{ km}, -4800\text{ km}, \ldots, 5000)$.

and prevailing winds are expected to broadly follow the latitudes, enhanced latitudinal dependence in an average sense (i.e., when, for example, many spatially lagged grids are considered in the averaging, or when, for example, large-scale and long-term dependence is computed) would appear to be intuitive. However, we would like to add a note of caution that the above intuition is valid only in an averaged sense and not at the scale of storms. In fact, since precipitation is produced by low-level convergence, a process that acts on scales of hundreds to thousands of kilometers to draw moisture into the system from all directions, the presence of strong cyclones/anticyclones is responsible for changes in the convergence, and not the latitudinal stratification of the winds. Correlation and tail dependence obtained from CCSM3 simulations for 1940–2005 display the latitudinal dependence as well (Fig. 13b). However, correlation and tail dependence based on the simulated data appear slightly tilted at an angle to the latitudinal structure, which suggests enhanced dependence, over and above the dependence based on geographical proximity, on the average along the longitudinal direction. The enhanced longitudinal dependence is on the average weaker than the latitudinal dependence given the degree of the tilt, however the tilt appears consistent for both correlation and tail dependence. While longitudinal dependence may be expected for isolated points or even at regional or local scales due to topographic or similar effects, a large-scale average dependence along longitudes

may be spurious. This appears especially true since the longitudinal component of the dependence is not apparent in the observed data (Fig. 13a).

By comparing observed and simulated data analyses for the period 1940–2005, higher values for correlation and tail dependence are observed from the simulated data as compared to that from the observed data (Fig. 13). From the observed data, no change is observed in correlation and tail dependence from 1965 to 2005 by considering two time windows 1965–1990 and 1980–2005 (plots not shown here). From CCSM3 simulations, the correlation and tail dependence plots for three time windows, i.e., 1965–1990, 1980–2005, and 2060–2099, are not different from that shown in Fig. 13b (plots not shown here).

4.4. Correlation and tail dependence between one specific grid and all grids in South America

The pairwise correlation and tail dependence between one specific grid and all other grids in South America are analyzed. One specific grid each in the Amazon basin, Brazilian Highlands, Mato Grasso Plateau, and São Paulo are chosen to investigate if precipitation and precipitation extremes occurring in these areas have any effects on precipitation and precipitation extremes in other parts of South America. Fig. 14 shows correlation and tail dependence corresponding to the grids in the Amazon Basin, Brazilian Highlands, Mato Grasso Plateau, and São Paulo.

In the Amazon basin, correlation between the chosen grid, i.e., Manaus in Brazil, and small areas around that grid, is high compared to the other parts of South America (Fig. 14a). This indicates that precipitation at Manaus is correlated with the areas surrounding it, but there is a very small amount of dependence between extremes at Manaus and its neighboring areas (Fig. 14a). The precipitation at one specific grid in the Brazilian Highlands is highly correlated with the precipitation at all grids in the Brazilian Highlands (Fig. 14b). The precipitation correlation is also high between one specific grid in the Brazilian Highland and some grids in highly populated coastal eastern regions of Brazil. The tail dependence at one specific grid and all other grids in the Brazilian Highlands is also high compared to the other parts of South America. In the Mato Grasso Plateau, correlation between one specific grid and all other grids in the plateau varies from 0.2 to 0.5 (Fig. 14c). The precipitation in the Mato Grasso Plateau is also correlated with some portions of the Brazilian highlands. The tail dependence is only confined to the vicinity of that grid and is very low at the other grids in the plateau. A very high correlation is observed between precipitation in São Paulo and highly populated coastal regions including Rio De Janeiro (Fig. 14d). Correlation is also high between one specific grid in São Paulo and some grids in the Mato Grasso Plateau. The tail dependence between São Paulo and some coastal cities close to São Paulo is also high compared to the other parts of South America.

From the analyses of two overlapping time windows 1965–1990 and 1980–2005, correlation between one specific grid in the Mato Grasso Plateau and its neighboring areas appears to exhibit an increasing trend (Fig. 15). No change is observed in the tail dependence from 1965 to 2005. Correlation and tail dependence between one specific grid in the Amazon basin, Brazilian Highlands, and São Paulo, with all other grids in South America, remain unchanged from 1965 to 2005 (results not shown here). We note that time windows are used here to identify trends in the dependence to a first order. However, overlapping time windows are necessary to provide adequate data within each window for estimation purposes.

Correlation and tail dependence between one specific grid in the Amazon basin, Brazilian Highlands, Mato Grasso Plateau, and São Paulo, with all other grids in South America, for the simulated data from CCSM3 are shown in (Fig. 16) from 1940 to 2005. The dependence, i.e., correlation and tail dependence, show more *regular* looking shapes with lesser spread for the simulations compared to the observations (Fig. 14 and Fig. 16). One reason may be that climate model simulations are based on partial differential equations which are expected to lead to smoother structures. When comparing dependence from observations and simulations in the Amazon basin, the observed correlation is found to be higher for the observed data whereas the tail dependence is not significantly different (Fig. 14a and Fig. 16a). In the Brazilian Highlands, the observations show greater correlation and tail dependence than the simulated data (Fig. 14b and Fig. 16b). The simulated data show greater correlation and tail dependence in the Mato Grasso Plateau compared to the observed data (Fig. 14c and Fig. 16c). No change in correlation and tail dependence is observed in São Paulo in either observations or simulations (Fig. 14d and Fig. 16d). The higher correlation and tail dependence around one specific grid in the Brazilian highlands, as well as in Sao Paulo, which are obtained consistently from both observed and simulated data, appear to correspond to the more varied orography in those regions.

The simulated data from CCSM3 are analyzed for three time periods, i.e., 1965–1990, 1980–2005, and 2060–2099, to find the temporal variability of correlation and tail dependence. The tail dependence in the Brazilian Highlands and correlation in the Mato Grasso Plateau increase from 1965 to 2005 (Fig. 17) whereas they are not different for 1980–2005 and 2060–2099. Correlation does not change in the Amazon basin, Brazilian Highlands, and São Paulo whereas the tail dependence remains same in the Amazon basin, Mato Grasso Plateau, and São Paulo for all three time windows (plots not shown here).

5. Summary and conclusions

An approach for quantifying and visualizing dependence among extreme values in space and time was presented. This study utilized weekly maxima residuals

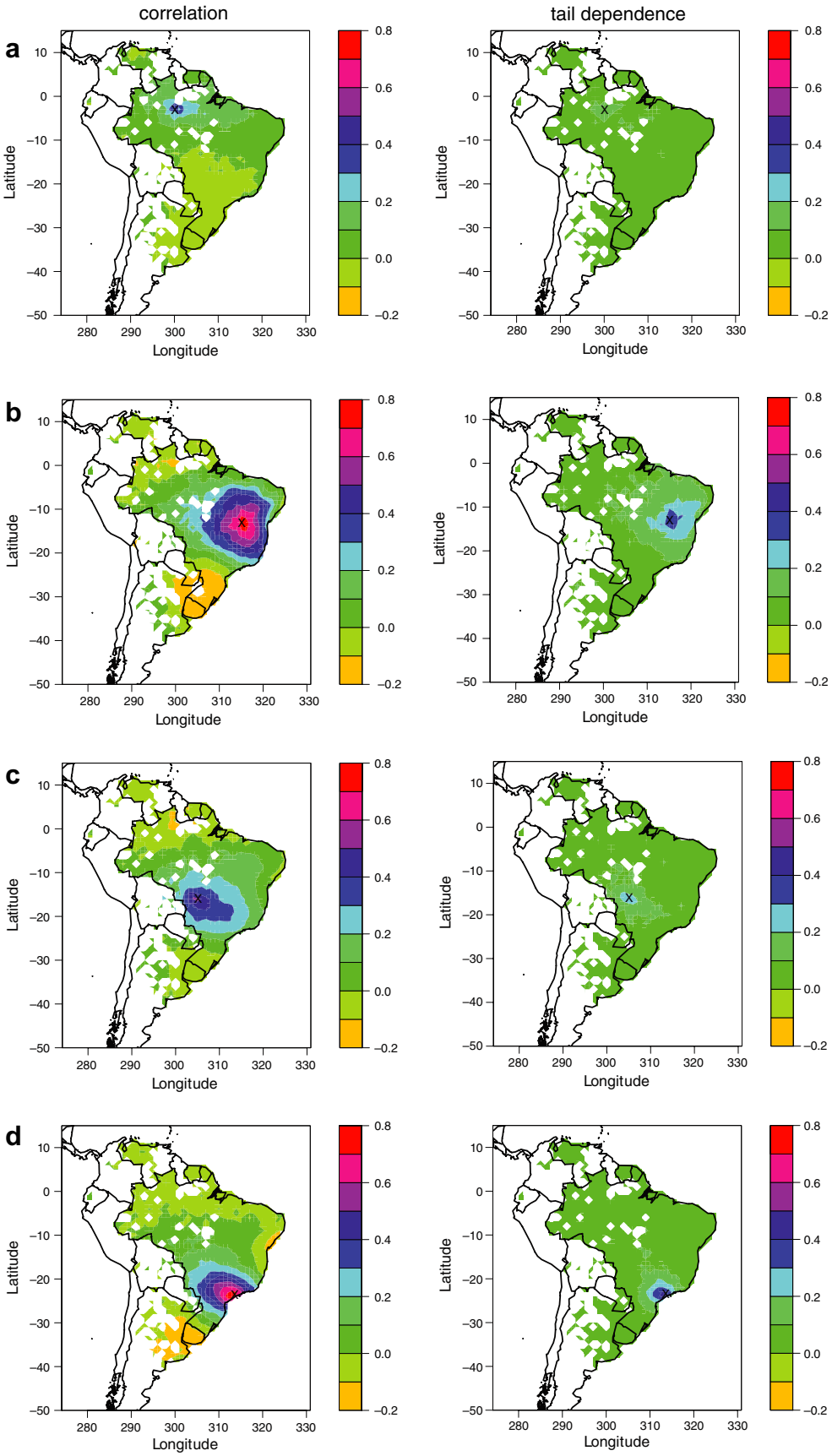

Fig. 14. Observed data: correlation and tail dependence between one specific grid (marked with a cross) in the (a) Amazon basin; (b) Brazilian highlands; (c) Mato Grasso Plateau; and (d) São Paulo and all grids in South America for the period 1940–2005. Left column: correlation. Right column: tail dependence.

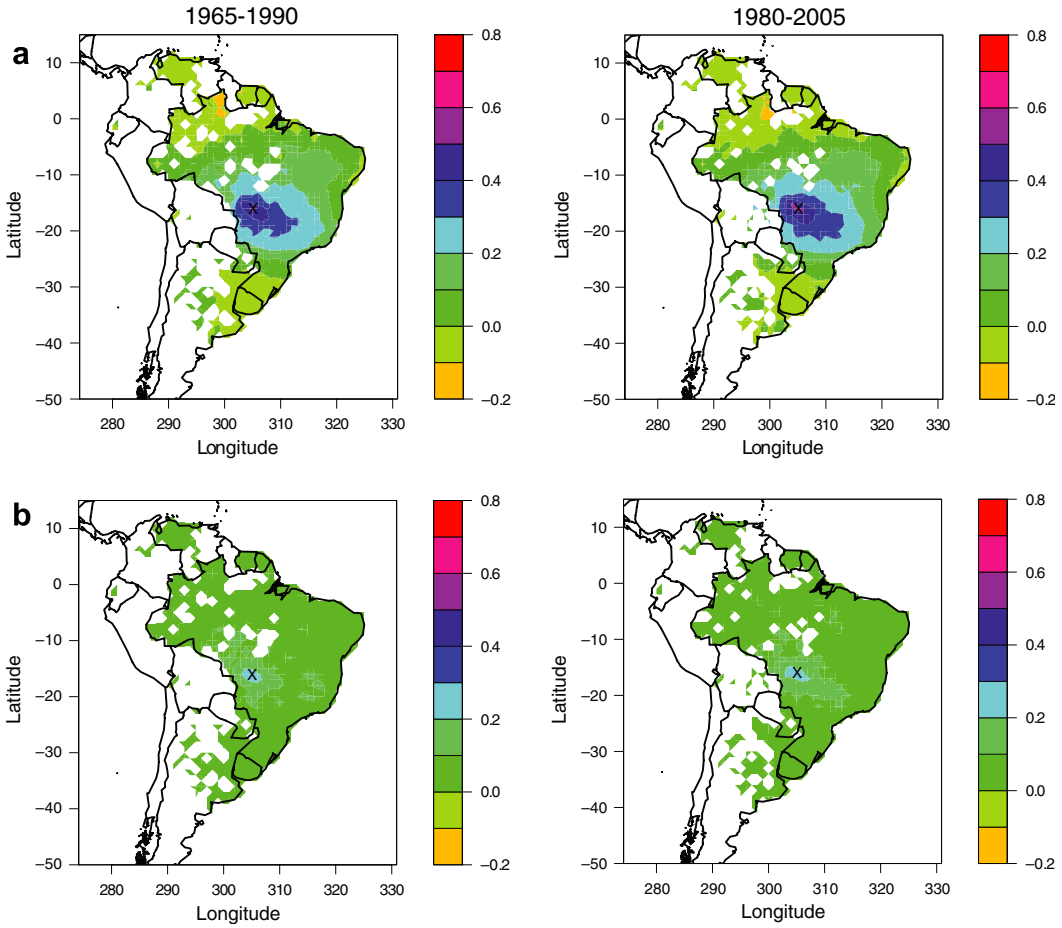

Fig. 15. Observed data: correlation and tail dependence between one specific grid (marked with a cross) in the Mato Grasso Plateau and all grids in South America for (a) correlation and (b) tail dependence. Left column: 1965–1990. Right column: 1980–2005.

precipitation from observations as well as climate simulations obtained from CCSM3 because they showed weaker temporal dependence compared to daily and weekly maxima precipitation. We analyzed observations and simulations for the period 1940–2005 to explore and quantify the correlation and tail dependence in space and time. The spatial and temporal variability of correlation and tail dependence between one specific grid in the Amazon basin, Brazilian Highlands, Mato Grasso Plateau and São Paulo, with all other grids in South America, were also investigated. In addition, two time windows, i.e., 1965–1990 and 1980–2005, were analyzed separately to understand the temporal change (or trends) in the spatial correlation and tail dependence. The results of our analysis were used to compare and contrast the correlation and tail dependence obtained from observations and climate model simulations for the period 1940–2005. Finally, climate simulations for the period 2060–2099 were investigated to understand and produce predictive scenarios for the correlation and tail dependence in the future. The important insights from this study are summarized below.

- Correlation and tail dependence obtained from the analysis of simulations from CCSM3 are higher as compared to that from observations when the distance between grids is greater than 1700 km and 1500 km, respectively. They appear to increase for observations from 1965 to 2005 based on two overlapping time windows, whereas the simulated data show no change in correlation and tail dependence for 1965–1990, 1980–2005, and 2060–2099.
- Greater correlation and tail dependence in the latitudinal direction are observed as compared to the longitudinal direction for both observations and simulations. The simulated data indicate more correlation and tail dependence as compared to that from the observed data, and appears to exhibit a slight tilt in the longitudinal direction.
- Correlation and tail dependence between one specific grid in the Amazon basin, Brazilian Highlands, Mato Grasso Plateau, and São Paulo, with all other grids in South America, show higher values over neighboring areas for both observations and simulations. For observations, the tail dependence around one specific grid in the Brazilian Highlands and São Paulo extends to some highly populated coastal eastern regions of Brazil. When compared to simulations, the observations show greater correlation around one specific grid in the Amazon

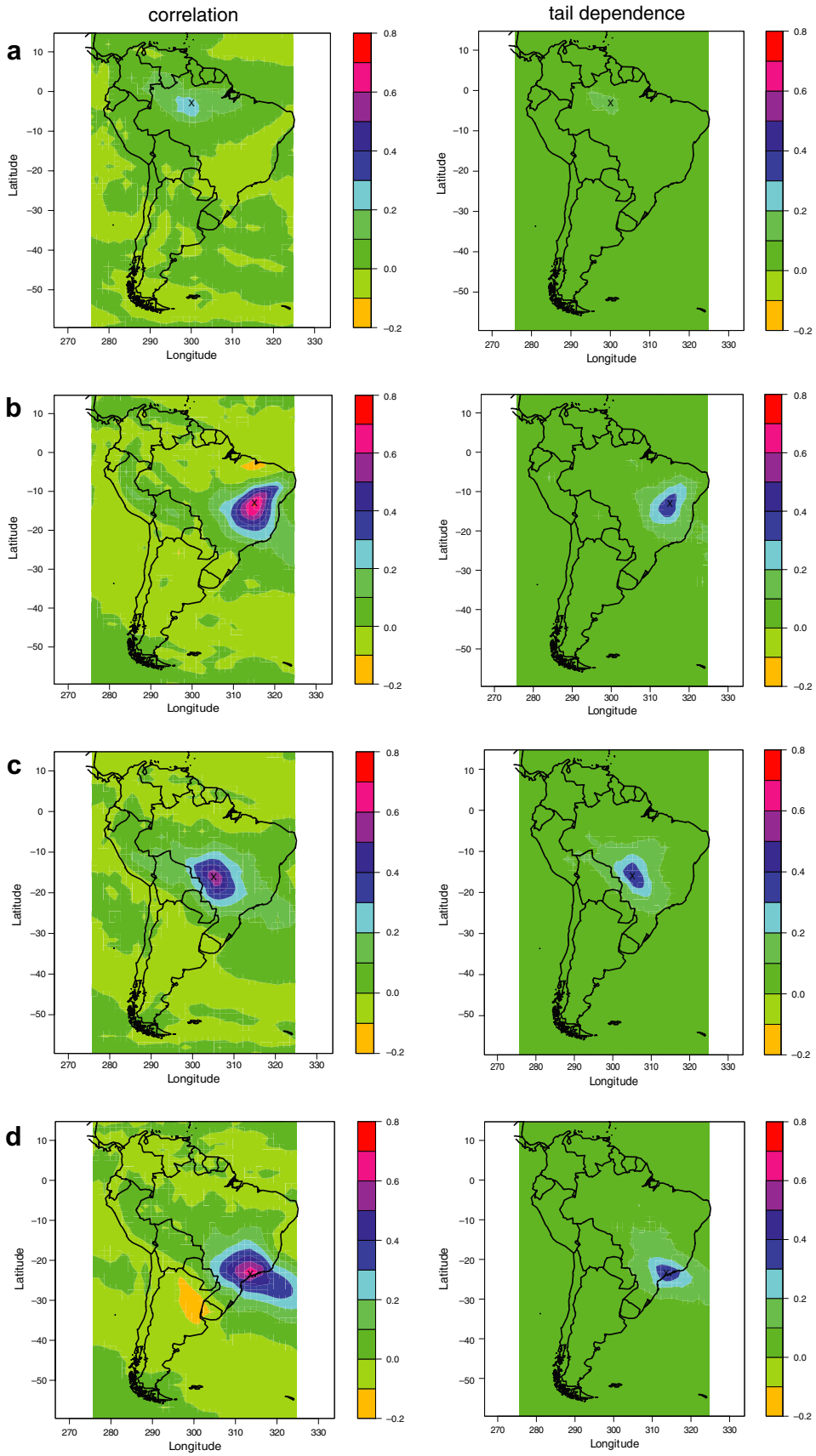

Fig. 16. Simulated data from the CCSM3 climate model: correlation and tail dependence between all grids in South America with one specific grid (marked with a cross) in the (a) Amazon basin; (b) Brazilian highlands; (c) Mato Grasso Plateau; and (d) São Paulo for the period 1940–2005. Left column: correlation. Right column: tail dependence.

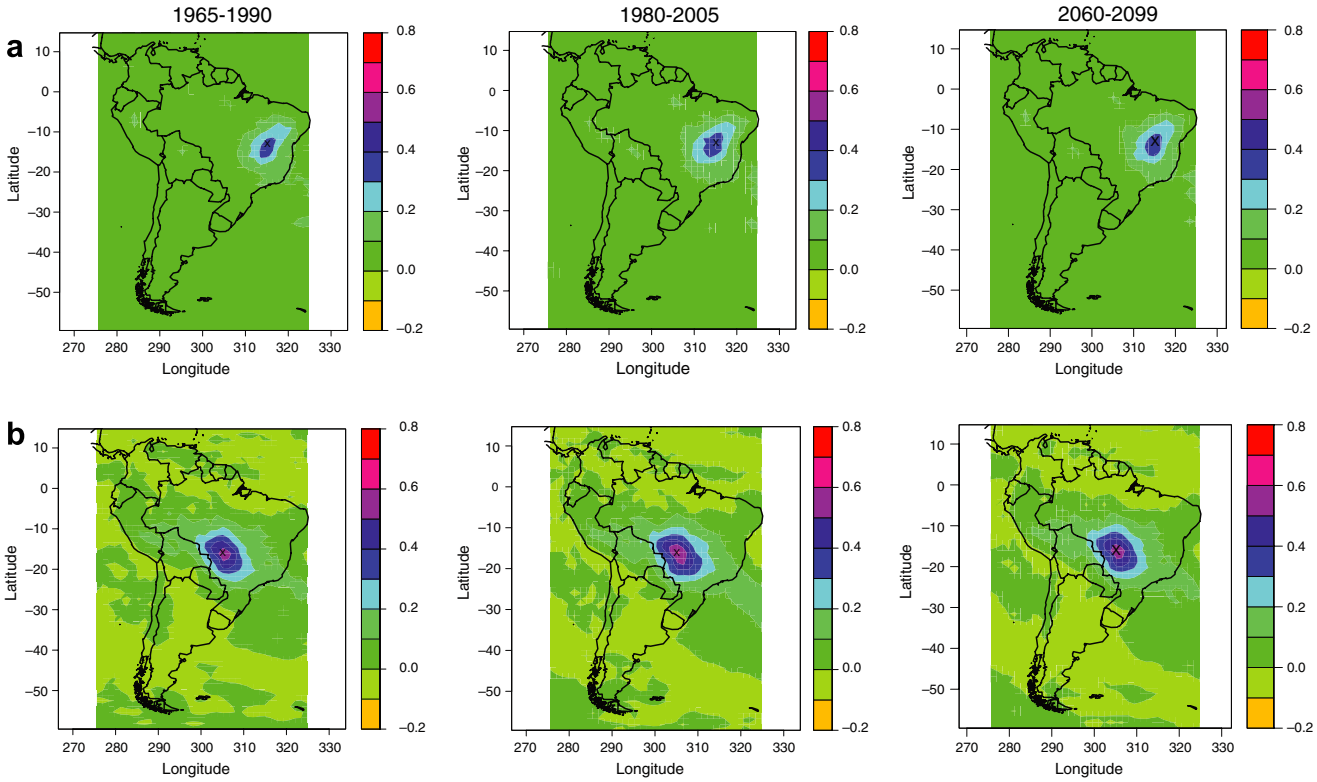

Fig. 17. Simulated data from the CCSM3 climate model: (a) tail dependence between one specific grid (marked with a cross) in the Brazilian Highlands and all grids in South America and (b) correlation between one specific grid in the Mato Grasso Plateau and all grids in South America. Left column: 1965–1990. Middle column: 1980–2005. Right column: 2060–2099.

basin and Brazilian Highlands and greater tail dependence in the Brazilian Highlands. The simulated data show greater correlation and tail dependence in the Mato Grasso Plateau as compared to that from the observed data. Correlation around one specific grid in the Mato Grasso Plateau increases for both observations and simulations from 1965 to 2005 based on two overlapping time windows. There is an increase in tail dependence in the Brazilian Highlands only in the case of simulated data from 1965 to 2005.

The results and interpretations of our analysis are valid at the space–time scales at which they were generated, specifically $1^0 \times 1^0$ spatial grids and weekly, and should not be extrapolated to higher resolutions, for example, station precipitation or daily data. The approaches for quantifying spatio-temporal dependence among the usual and the extreme values, developed in this study, can have significant impacts in the analysis of geographic and/or spatio-temporal data in multiple domains. These include, but are not limited to, the analysis of remotely sensed data from satellites and aerial photography as well as data collected from in situ wireless sensor networks and large-scale sensor infrastructures. In addition, they can be used to study the inter-dependence among precipitation and temperature extremes. Such studies can have significant impacts on our understanding of crop yields as well as extremal teleconnections in climate or hydrology.

While mathematical simulations, especially in complex phenomena like hydrology and climate, may be able to capture trends, they are known to be inadequate for extremes because such models tend to capture the predominant or usual behavior rather than the unusual. In addition, climate models are often tested by their ability to generate realistic estimates of precipitation, as the latter are known to be difficult to simulate owing to thresholds and intermittencies, possible dominance of sub-grid scale features like convection, as well as the fact that precipitation is a derived quantity and not a state variable of the climate models. Thus, a comparison of observed and climate model simulated precipitation extremes can be a hard test for climate models. However, when extremes are of interest, such comparisons are important, and can lead directly to uncertainty estimates for the corresponding climate model projections in the future. This study not only generates insights regarding the dependence among extreme precipitation values in space and time, but allows for a thorough comparison between observations and model simulations in terms of this dependence. The comparison is performed both in an aggregate sense, e.g., when the average dependence at multiple spatial lags is determined, as well as in more detail when the dependence between one selected point and all other points are determined. This can lead to better understanding and quantification of the uncertainty in the climate model simulations. Future research needs to extend this study not only to compare

the dependence between extremes but also the parameters of the extreme value distributions, as well as compute and visualize the uncertainties. The ultimate goal would be to generate projections in the future regarding the properties of extremes, as well as estimates of the uncertainties surrounding the projections.

Future research will investigate the dependence among multiple hydrological and climatic variables, e.g., one important area would be to explore the spatio-temporal dependence among heat waves or cold spells, i.e., the geospatial–temporal extremes of temperature. Future improvements to the methodologies include rigorous approaches for bias removal in the tail dependence measure, incorporation of sophisticated approaches for extracting the temporal dependence structures and the development of uncertainty quantification formulations that can relate the exceedence of the re-scaled test statistic proposed earlier over unity to the overall uncertainty estimation. New methodological extensions can be developed to explore the possibility of allowing extremal parameters, or extremal dependence parameters, to be functions of space and time. These extensions have been proposed by previous researchers for univariate extreme value theory, in the context of temporal trends in the extremes as well as physically based covariates [3–6]. However, extensions to multivariate extremal dependence may be rather involved, especially if the dependence changes with space and time. Finally, the eventual goal would be to relate the insights obtained from this and similar studies to improvements in our understanding of the precipitation and climate physics at multiple scales. Since precipitation extremes may be linked to El Niño-Southern Oscillation (ENSO) [53,54] and Madden–Julian Oscillation (MJO) [55], we will investigate the influence of ENSO and MJO on the spatial dependence among precipitation extremes. Future research will also tie this effort with our previous work on nonlinear dependence between ENSO and the tropical hydrological cycle [56].

Acknowledgements

This research was funded by the SEED money funds of the Laboratory Directed Research and Development Program of the Oak Ridge National Laboratory (ORNL), managed by UT-Battelle, LLC for the US Department of Energy (DOE) under Contract No. DE-AC05-00OR22725. We would like to gratefully acknowledge Dave Allured and Brant Liebmann from NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado for providing us the precipitation observations in South America. We acknowledge a couple of modeling groups for providing us the CCSM3 simulations, namely, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the model output, and the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) for organizing the model data analysis activity. The multi-model data archive is supported by the Office of Science, US DOE. The authors are thankful to three anonymous reviewers for their helpful suggestions which significantly improved the quality of the paper. Shiraj Khan and Auroop Ganguly would like to acknowledge the help and support provided by Prof. Sunil Saigal of Civil and Environmental Engineering at the University of South Florida. Auroop Ganguly would like to thank Drs. David J. Erickson III and George Ostrouchov of ORNL, as well as Dr. Rick Katz of the National Center for Atmospheric Research and Prof. Tailen Hsing of the Ohio State University, for their collaboration, support and encouragement on the SEED project.

References

[1] Kuhn G. On dependence and extremes. PhD thesis (Advisor: C. Klüppelberg), Munich University of Technology, 2006.
[2] Coles SG, Tawn J. A Statistical methods for multivariate extremes: an applications to structural design. Appl Statist 1994;43(1):1–48.
[3] Coles SG. An introduction to statistical modeling of extreme values, Springer Series in Statistics. London: Springer; 2001.
[4] Katz R, Parlange M, Naveau P. Statistics of extremes in hydrology. Adv Water Resour 2002;25:1287–304.
[5] Engeland K, Hisdal H, Frigessi A. Practical extreme value modelling of hydrological floods and droughts: a case study. Extremes 2004;7:5–30.
[6] Katz R, Brush G, Parlange M. Statistics of extremes: modeling ecological disturbances. Ecology 2005;86:1124–34.
[7] Casson E, Coles SG. Spatial regression models for extremes. Extremes 1999;1:449–68.
[8] Begueŕia S, Vicente-Serrano S. Mapping the hazard of extreme rainfall by peaks over threshold extreme value analysis and spatial regression techniques. J Appl Meteorol Climatol 2006;45:108–24.
[9] Coles SG, Heffernan JE, Tawn JA. Dependence measures for extreme value analysis. Extremes 1999;2:339–65.
[10] Ledford AW, Tawn JA. Diagnostics for dependence within time series extremes. J R Statist Soc Ser 2003;B 65(2):521–43.
[11] Heffernan JE, Tawn JA. A conditional approach for multivariate extreme values. J R Statist Soc Ser 2004;B 66(3):497–546.
[12] Groisman PY, Karl TR, Easterling DR, Knight RW, Jamason PF, Hennessy KJ, et al. Changes in the probability of heavy precipitation important indicators of climatic change. Climatic Change 1999;42:243–83.
[13] Haylock M, Goodess C. Interannual variability of European extreme winter rainfall and links with mean large-scale circulation. Int J Climatol 2004;24:759–76.
[14] Coles SG, Tawn JA. Modelling extremes of the areal rainfall process. J R Statist Soc 1996;B 58(2):329–47.
[15] Schlather M, Tawn JA. A dependence measure for multivariate and spatial extreme values: properties and inference. Biometrika 2003;90(1):139–56.
[16] Liebmann B, Jones C, Carvalho LMV. Interannual variability of daily extreme precipitation events in the state of São Paulo, Brazil. J Clim 2001;14:208–18.
[17] Carvalho LMV, Jones C, Liebmann B. Extreme precipitation events in southeastern South America and large-scale convective patterns in the South Atlantic Convergence Zone. J Clim 2002;15:2377–94.
[18] Easterling DR, Evans JL, Groisman PY, Karl TR, Kunkel KE, Ambenje P. Observed variability and trends in extreme climate events. Bull Am Met Soc 2000;81:417–25.
[19] Meehl GA, Tebaldi C. More intense, more frequent, and longer lasting heat waves in the 21st century. Science 2004;305:994–7.
[20] Liebmann B, Allured D. Daily precipitation grids for South America. Bull Am Meteor Soc 2005;86:1567–70.

[21] Goswami BN, Venugopal V, Sengupta D, Madhusoodanan MS, Xavier PK. Increasing trends of extreme rain events over India in a warming environment. Science 2006;314:1442.
[22] Collins WD, Bitz CM, Blackmon ML, Bonan GB, Bretherton CS, Carton JA, et al. The community climate system model version 3 (CCSM3). J Clim 2006;19(11):2122–43.
[23] Branstetter ML, Erickson III DJ. Continental runoff dynamics on the community climate system model version 2 (CCSM2) control simulation. J Geophys Res 2003;108(D17):4550. doi:10.1029/2002JD003213.
[24] Meehl GA, Arblaster JM, Tebaldi C. Understanding future patterns of increased precipitation intensity in climate model simulations. Geophys Res Lett 2005;32. doi:10.1029/2005GL023680.
[25] Boville BA, Rasch PJ, Hack JJ, McCaa JR. Representation of clouds and precipitation processes in the community atmosphere model version 3 (CAM3). J Clim 2006;19(11):2184–98.
[26] Gaines SD, Denny MW. The largest, smallest, highest, lowest, longest, and shortest: extremes in ecology. Ecology 1993;74:1677–92.
[27] Kunkel KE, Changnon SA, Shealy RT. Temporal and spatial characteristics of heavy-precipitation events in the Midwest. Mon Wea Rev 1993;121:858–66.
[28] Yates D, Gangopadhyay S, Rajagopalan B, Strzepek K. A technique for generating regional climate scenarios using a nearest-neighbor algorithm. Water Resour Res 2003;39(7):1199. doi:10.1029/2002WR001769.
[29] Michael JR. The stabilized probability plot. Biometrika 1983;70(1):11–7.
[30] Coles SG. On goodness-of-fit tests for the two-parameter Weibull distribution derived from the stabilized probability plot. Biometrika 1989;76(3):593–8.
[31] Kimber AC. Tests for the exponential, Weibull and Gumbel distributions based on the stabilized probability plot. Biometrika 1985;72(3):661–3.
[32] Davison A, Smith R. Models for exceedances over high thresholds (with discussion). J R Statist Soc Ser 1990;B 52:393–442.
[33] Smith R, Robinson P. A bayesian approach to the modelling of spatial–temporal precipitation data. In: Gatsonis C, editor. Case studies in bayesian statistics III. Lecture notes in statistics 121, 121. New York: Springer; 1997. p. 237–64.
[34] Smith R, Tawn JA, Coles SG. Markov chain models for threshold exceedances. Biometrika 1997;84:249–68.
[35] Smith R. Statistics of extremes, with applications in environment, insurance and finance. In: Finkenstädt B, Rootzén H, editors. Extreme values in finance, telecommunications and the environment. Chapman and Hall/CRC Press; 2003. p. 1–78 [chapter 1].
[36] Yun S, Smith R. Spatial trends and spatial extremes in South Korean ozone. J Korean Statist Soc 2003;32(4):313–35.
[37] Kendall M, Gibbons J. Rank correlation methods. 5th ed. London: Edward Arnold; 1990.
[38] Joe H. Multivariate models and dependence concepts. Monographs on statistics and applied probability, vol. 73. London: Chapman & Hall; 1997.
[39] Nelsen R. An introduction to copulas. Lecture notes in statistics, Vol. 139. New York: Springer; 1999.
[40] Hult H, Lindskog F. Multivariate extremes, aggregation and dependence in elliptical distributions. Adv Appl Probab 2002;34(3):587–608.
[41] Embrechts P, McNeil A, Straumann D. Correlation and dependence in risk management: properties and pitfalls. Risk management: value at risk and beyond (Cambridge, 1998). Cambridge: Springer; 2002. p. 176–223.
[42] Lindskog F. Linear correlation estimation, preprint, ETH Zurich. <citeseer.ist.psu.edu/lindskog00linear.html> 2000.
[43] Resnick S. Extreme values, regular variation, and point processes. New York: Springer; 1987.
[44] Hsing T, Klüppelberg C, Kuhn G. Dependence estimation and visualization in multivariate extremes with applications to financial data. Extremes 2004;7(2):99–121.
[45] Einmahl J, de Haan L, Piterbarg V. Nonparametric estimation of the spectral measure of an extreme value distribution. Ann Statist 2001;29(5):1401–23.
[46] Embrechts P, Klüppelberg C, Mikosch T. Modelling extreme events for insurance and finance. Berlin, Heidelberg: Springer; 1997.
[47] Tawn JA. Bivariate extreme value theory: models and estimation. Biometrika 1988;75(3):397–415.
[48] Ledford AW, Tawn JA. Modelling dependence within joint tail regions. J R Statist Soc Ser 1997;B 59(2):475–99.
[49] Peng L. Second order condition and extreme value theory. PhD thesis (Advisor: L. de Haan), Erasmus University, Rotterdam, 1998.
[50] Einmahl JHJ, Li J, Liu RY. Extreme value theory approach to simultaneous monitoring and thresholding of multiple risk indicators. CentER Discussion Paper (Int. rep. 2006-104) Econometrics. 2006.
[51] Frahm G, Junker M, Schmidt R. Estimating the tail dependence coefficient. Insurance: Math Econ 2005;37:80–100.
[52] Ledford AW, Tawn JA. Statistics for near independence in multivariate extreme values. Biometrika 1996;83(1):169–87.
[53] Gershunov A, Barnett TP. ENSO influence on intraseasonal extreme rainfall and temperature frequencies in the contiguous United States: observations and model results. J Clim 1998;11:1575–86.
[54] Kane RP. Rainfall extremes in some selected parts of Central and South America: ENSO and other relationships reexamined. Int J Climatol 1999;19:423–55.
[55] Jones C, Waliser DE, Lau KM, Stern W. Global occurrences of extreme precipitation and the Madden–Julian oscillation: observations and predictability. J Clim 2004;17:4575–89.
[56] Khan S, Ganguly AR, Bandyopadhyay S, Saigal S, Erickson III DJ, Protopopescu V, et al. Nonlinear statistics reveals stronger ties between ENSO and the tropical hydrological cycle. Geophys Res Lett 2006;33:L24402. doi:10.1029/2006GL027941.